수산경영론집, 제39권 제3호 2008. 12

Developing an Evaluation Tool of RFID-based Traceability Systems†

Kim, Jin-Baek*

〈 Contents 〉

Ⅰ. Introduction
Ⅱ. The Previous Research and Measure Derivation
  1. Food Risk
  2. Network
  3. Information System
  4. RFID
Ⅲ. The Tool Development Process
  1. Formulation of a Tentative Evaluation Tool
  2. Data Collection and Sample Homogeneity
  3. Refinement of the Evaluation Tool
  4. Validation of the Evaluation Tool
  5. Relative Importance of the Constructs
Ⅳ. Conclusions
Appendix
References
Abstract

Ⅰ. Introduction

Many information technologies are applied in food sectors to improve food quality and safety. Among them, traceability systems are introduced for food safety (Houghton et al., 2008; Miraglia et al., 2004: Seino et al., 2004; Smith et al., 2005). To identify the causes of food risk, all participants in food supply chain should share their own production or handling information through traceability systems. Therefore, a traceability system is implemented as an inter-orga-

접수 : 2008년 9월 17일　　최종심사 : 2008년 10월 14일　　게재확정 : 2008년 10월 21일
† This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-041-b00180).
* College of Business Administration, Tongmyong University(Corresponding author: 051-629-1864, jinkim@tu.ac.kr)

nizational information system on the basis of network technology. To trace food products efficiently, traceability systems also need a capability of identifying individual food products. To obtain the capability, traceability systems began to adopt RFID technology instead of bar code technology (Kärkkäinen & Holmström, 2002).

Although the implementation of RFID – based traceability systems (RFID – TS) has been increasing in food sectors (Haapala, 2003; Kärkkäinen, 2003; Najjar et al., 1997; Wang et al., 2006), it is still unknown to what extent and how well the developed RFID – TS works. Therefore, its evaluation tool is required to measure the performance of RFID – TS (Hou & Huang, 2006; ISO, 1999; Van Der Spiegel et al., 2004). To develop an evaluation tool, its performance aspects should be focused on in advance. Unfortunately, there have been few studies on RFID – TS performance. To accomplish its goal as a risk reliever, RFID – TS is required to be equipped with three main components: RFID, network, and information system (Taniguchi & Sagawa, 2005). Therefore, as shown in Figure 1, this study adopted its goal and main components as performance aspects of RFID – TS.

To develop an evaluation tool of RFID – TS, this study followed Churchill (1979)' s paradigm[1] that was utilized in many studies (Koste et al., 2004; Parasuraman et al., 1988; Perrin – Martinenq, 2004). In accordance with

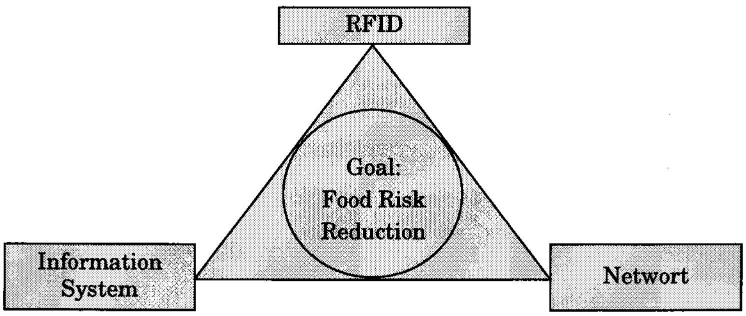

Figure 1. Performance Aspects of RFID – TS

1) Churchill's paradigm is a framework involving an iterative process. The framework is appropriate when the constructs are multi-dimensional and complex. It consists of 2 stages. At the first stage, we must specify the domain of the construct and then generate a sample of items for its measurement. Next is a process of data collection and purification of the measure based on factor analysis and the calculation of coefficient alpha. At the second stage, another data set is used. After a new data set is collected, an assessment of the reliability and validity of the construct is performed.

Churchill' s paradigm, literature reviews were firstly done to derive measures. The derived measures were examined by a focus group and formulated as a tentative tool. The tentative tool was refined through two stages using their own datasets. Finally, the structural stability and construct validity of the refined tool were assessed by confirmatory factor analysis and AVE test.

Ⅱ. The Previous Research and Measure Derivation

1. Food Risk

The concept of risk perception has been explored by many researchers. Roselius (1971) used a perceived risk concept to measure attitudes about the helpfulness of risk relievers. To identify which risk reliever was more effective in reducing purchase related risks, he used four risk measures: time, hazard, money, and ego risks.

Kaplan et al. (1974) added performance risk to Roselius (1971)' framework, but excluded time risk and separated ego risk into psychological and social risks. Mitchell & Greatorex (1988) used financial, functional, physical and social risks in the study of consumer risk perception. Mitra et al. (1999) measured purchase related risks with Kaplan et al. (1974)' s framework. But they included an additional risk measure, time risk, which was identified by Roselius (1971). Yeung & Morris (2001) used the concept of risk perception to develop a conceptual model of consumer food purchase. They included six risk measures: financial, psychological, social, performance, physical, and time risks. Mahon & Cowan (2004) used Yeung & Morris (2001)' s framework to measure perceived risk in purchasing food.

According to Churchill' s paradigm, this study used a focus group to screen the initial measures. The screening objectives were to avoid conceptual redundancies among the measures and delete some unsuitable measures. The focus group consisted of three experts majoring in information management or fisheries management. Depending on the examination of the focus group, functional risk in Mitchell & Greatorex (1988) was similar to performance risk used in other studies. And almost studies tended to use psychological and social risks instead of ego risk. Therefore, it was concluded that many studies used six risk measures: financial, psychological, social, performance, physical, and time risks. These six risk measures were grouped into two categories; one was economic risk and another was non-economic risk. The economic risk category included financial, performance, and time risks. The non-economic risk category included psychological, social, and physical risks.

2. Network

Rodrigues (1990) used four measures to evaluate network performance including efficiency, response time, throughput, and fairness. Throughput is the amount of work done in a given time period. So throughput is similar to workload.

French et al. (1990) adopted several measures to compute network performance such as connectivity, throughput, delay, and grade of service. Connectivity is the fraction of node pairs remaining connected in a damaged network. Grade of service is the average probability that node pairs are connected. Therefore, connectivity is similar to reliability and grade of service is similar to usage.

Bayrak & Grabowski (2003) derived measures from mathematical models, large-scale system models, and statistical and technical communication models. After examining the various models, they selected five network performance measures: reliability, correctness, response time, workload, and usage.

To evaluate network service quality, Zhang & Mukherjee (2004) discussed some measures such as availability, reliability, restoration time, and restorability. These measures add up to robustness. Therefore, all these measures are related to the reliability measure of Bayrak & Grabowski (2003).

Table 1. Network Measures

Researchers	RE	CO	RT	WL	UA	SE	IN	EF	FA
Rodrigues (1990)			○	△				○	○
French et al. (1990)	△		△	△	△				
Bayrak & Grabowski (2003)	○	○	○	○	○				
Zhang & Mukherjee (2004)	△								
Lo Bello et al. (2004)	△		△			○	○		

Note) RE: reliability, CO: correctness, RT: response time, WL: workload, UA: usage, SE: security, IN: interface, EF: efficiency, FA: fairness; ○ same word, △ similar word.

Lo Bello et al. (2004) regarded a traceability system as a collaborative information system. They evaluated network performance with some measures such as robustness, security, user interface, and access time.

Depending on the examination of the focus group, it was concluded that some network measures were similar to one another. That is, it was ascertained that some measures were simply named differently. For example, access time in Lo Bello et al. (2004) and delay in French et al. (1990) were conceptually similar to response time in Rodrigues (1990). Also availability, restoration time, and restorability in Zhang & Mukherjee (2004), connectivity in French et al. (1990), and robustness in Lo Bello et al. (2004) were conceptually similar to reliability in Bayrak & Grabowski (2003). Therefore, to avoid conceptual redundancies among network measures, nine distinct network measures were selected as a result of the focus group' s examination. As shown in Table 1, nine network measures were reliability, correctness, response time, workload, usage, security, interface, efficiency, and fairness.

3. Information System

DeLone & McLean (1992) have greatly influenced the information system (IS)

Table 2. IS Measures

Researchers	System	Information	Service
DeLone & McLean (2004)	Scalability, Interactivity	Accuracy, Relevance, Understandability, Completeness, Currency, Competitive intelligence	Responsiveness, Assurance, Empathy, Following-up
	N/W: Usability, Download time, Responsiveness, Reliability, Flexibility, Usefulness, Security		
Chang & King (2005)	Effect on job, Effect on external constituencies, Effect on internal processes, Effect on knowledge and learning	Presentational quality, Accessibility, Flexibility	Intrinsic quality of provider, Interpersonal quality of provider, IS training, Flexibility
	N/W: System usage quality, Intrinsic system quality	*D & M: Intrinsic quality, Reliability, Contextual quality, Usefulness*	D & M: *Responsiveness*
Golan et al. (2004)		Breadth, Depth, Precision	

Note) N/W: redundant measures with network measures, D & M: similar measures to DeLone & McLean (2004)' s measures performance area. Since their paper was published, many researchers (Igbaria & Tan, 1997; Rai et al., 2002; Seddon & Kiew, 1994) have made use of their model. DeLone & McLean (2004) suggested an updated version of their existing model. They proposed two categories of performance measures; one category included traditional IS measures and another category included e-commerce measures. Traditional IS measures had three dimensions: system, information, and service. As shown in Table 2, their three dimensions included several specific measures.

Chang & King (2005) proposed IS measures from an overall functional perspective. They used three output related dimensions: system, information, and service referring to a modified IO model, Pitt et al. (1995)' s model, and DeLone & McLean (2003)' s model. But the specific measures of each construct were somewhat different with DeLone & McLean (2004)' s measures.

Golan et al. (2004) focused on only information dimension. They insisted that a traceability system for tracking every input and process would be enormous and very costly. Therefore, for RFID-TS to be cost-effective, it is necessary to determine the breadth, depth, and precision of information. Breadth is related to the amount of information collected. Depth is how far back or forward the system tracks. Precision is related to the degree of assurance with which a traceability system can pinpoint a particular product' s movement or characteristics.

Depending on the examination of the focus group, many IS measures were redundant with network measures or mutually similar. In system dimension, some measures were conceptually redundant with network measures such as reliability, usage, response time, and security. Those redundant measures were usability, download time, responsiveness, reliability, flexibility, usefulness, and security in DeLone & McLean (2004), and system usage quality and intrinsic system quality in Chang & King (2005). And in information dimension, intrinsic quality, reliability, and contextual quality in Chang & King (2005) were conceptually similar to some measures in DeLone & McLean (2004). Intrinsic quality and reliability in Chang & King (2005) were partially similar to accuracy in DeLone & McLean (2004). Contextual quality in Chang & King (2005) was similar to several measures such as relevance, completeness, and currency in DeLone & McLean (2004). Finally, usefulness in Chang & King (2005) was partially similar to competitive intelligence in DeLone & McLean (2004). The similarity between DeLone & McLean's measure and Chang & King's measure was caused by the fact that Chang & King (2005) referred to DeLone & McLean (2003) which was another updated version of DeLone & McLean (1992). Therefore, only the remaining measures, excluding the conceptually redundant or similar measures, were used as IS measures.

4. RFID

Electronic identification means is a key technology for RFID-TS. The RFID technology has been drastically developed and improved for decades. It is replacing barcodes in the automatic identification area (Hou & Huang, 2006). RFID performance measures can be classified into two categories: management performance and technological performance measures. Technological performance measures were previously reviewed in the IS and network aspects. Here, it was tried to identify management performance measures from the RFID technology.

Sahin et al. (2002) focused on management performance improvements of RFID. They applied two criteria to evaluate the performance of RFID-TS. One was an automatic item identification capability. The other was an item-level identification capability. They insisted that the automatic item identification capability of RFID permitted many management performance improvements such as a reduction in labor costs, an increase in store selling area, an acceleration of physical flows, a shrinkage reduction, an efficient control of the supply chain, a knowledge of customer behavior, a knowledge of out of stock products, and a reduction of delivery disputes. They also insisted that barcode identified products at the SKU (stock keeping unit) level, but RFID identified products at an individual item level. This RFID capability permitted much improvement in the management of perishable items, returns management, tracing quality problems, recall management, and counterfeiting.

There were some other studies on the performance of RFID systems. Flores et al. (2005) studied the performance of RFID technology. Their study was very technically-oriented and focused on very low level measures. Brown & Russell (2007) and Reynolds & Lynch (2005) investigated critical success factors in RFID adoption. Bellotti et al. (2002) and Scholtz & Consolvo (2004) evaluated ubiquitous computing systems. By the focus group, these studies were not recognized as RFID' s management performance related studies because their measures were too technically-oriented or unsuitable to RFID' s management performance evaluation. Therefore, to evaluate management performance aspect of RFID technology, this study referred mainly to Sahin et al. (2002).

Ⅲ. The Tool Development Process

1. Formulation of a Tentative Evaluation Tool

The purpose of this study is the development of a multiple-item evaluation tool for RFID-TS performance. To develop the tool, Churchill' s paradigm was adopted because it was appropriate when the tool was multi-dimensional and complex Koste et al. (2004). According to Churchill' s paradigm, this study specified performance constructs and then derived the measures from previous researches. As previously indicated, some measures of the constructs were conceptually redundant or unsuitable. To resolve these problems, a focus group was utilized in the measure derivation step. The focus group was asked to

Table 3. A tentative evaluation tool

Construct	Sub-construct	Measure
Risk	Economic risk	Financial risk, Performance risk, Time risk
	Non-economic risk	Psychological risk, Social risk, Physical risk
RFID	Automatic identification	Labor costs, Store selling area, Physical flows, Shrinkage, Control of the supply chain, Knowledge of customer behavior, Knowledge of out of stock, Delivery disputes
	Item-level identification	Management of perishable items, Returns management, Tracing quality problems, Recall management, Counterfeiting
IS	System	Scalability, Interactivity, Effect on job, Effect on external constituencies, Effect on internal processes, Effect on knowledge and learning
	Information	Accuracy, Relevance, Understandability, Completeness, Currency, Competitive intelligence, Presentational quality, Accessibility, Flexibility, Breadth, Depth, Precision
	Service	Responsiveness, Assurance, Empathy, Following-up, Intrinsic quality of provider, Interpersonal quality of provider, IS training, Flexibility
Network	Stability	Reliability, Correctness, Security
	Usability	Response time, Workload, Usage, Interface, Efficiency, Fairness

examine and remove the improper or redundant measures for evaluating RFID－TS performance.

Depending on the examination results of the focus group, a tentative evaluation tool was formulated. As shown in Table 3, the tentative evaluation tool consisted of four constructs. Each construct had two or three sub－constructs consisting of several measures. Each measure was developed into multiple items in the questionnaire and was rated using a 7－point Likert scale, anchored by strongly disagree (1) at one end, to strongly agree (7) at the other.

2. Data Collection and Sample Homogeneity

Depending on Churchill' s paradigm, two datasets collected at two points in time are needed. But in this study, the datasets were collected at one point in time because the examination of the measures by the focus group had been already done. The collected questionnaires were randomly divided into two datasets. Each dataset consisted of 200 questionnaires as recommended by Parasuraman et al. (1988).

A traceability system is an important information system in the food industry. Therefore, most questionnaires(90.5%) were collected from the workers in the fisheries sector. They were fisheries officials or researchers (40.0%), workers of fisheries cooperatives (32.8%), and other fisheries related workers (17.8%). Fisheries officials or researchers belonged to the Korean Ministry of Maritime Affairs and Fisheries, the National Fisheries Products Quality Inspection Service of Korea, the National Fisheries Research and Development Institute of Korea, and the Korea Maritime Institute. Workers of fisheries cooperatives belonged to the National Federation of Fisheries Cooperatives of Korea and its member

Table 4. Test for sample homogeneity

Dependent variable	Levene's test		Mean	Standard deviation	t value	Significant level
	F value	Significant level				
Preference	0.355	0.552	5.39[a] 5.26[b]	1.120[a] 1.187[b]	1.119	0.264
Overall performance	0.667	0.415	5.15[a] 5.07[b]	1.066[a] 1.016[b]	0.912	0.362

Note) a: dataset 1's value, b: dataset 2's value cooperatives. And other fisheries related workers belonged to aquaculture, feed, trading, and seafood manufacture companies. And male respondents accounted for 74.5% of the samples.

Before refining the tentative evaluation tool, two tests for homoscedasticity and homogeneity of the samples was performed first. Two dependent variables, the preference and the overall performance to RFID – TS, were used for the tests. In this study, Levene' s test was used for the homoscedasticity test, and the results were shown in Table 4. Because the significances of F values (0.355, 0.667) were 0.552 and 0.415 respectively, it was concluded that there were no statistically significant differences in two datasets. The results of t – test for the sample homogeneity also showed that there were no statistically significant differences between each variable' s means by datasets. Based on the results of Levene' s tests and t – tests, it was estimated that the samples were collected from the same population.

3. Refinement of the Evaluation Tool

1) The First Refinement Stage Using Dataset 1

The refinement stage of the tentative evaluation tool began with reliability analysis in accordance with Churchill' s paradigm. Because of the multidimensionality of the constructs, reliability analysis was performed separately for nine sub – constructs and was done by Cronbach' s alpha. Only the items where the Cronbach' s alpha coefficients exceeded 0.7 were accepted for exploratory factor analysis (EFA). Reliability analysis and EFA were performed iteratively. When the tentative evaluation tool was subjected to oblique rotation (oblimin), no clear factor pattern emerged. It was caused by the fact that several items had high loadings on more than one factor. In consequence, some constructs were deleted or meaningless. Therefore, EFA was accomplished by an orthogonal rotation method (varimax).

The iterations of reliability analysis and EFA were repeated twice. The result of the second iteration was presented in the Appendix entitled with "The first refined result of the evaluation tool."Nine measures were deleted through the iterations. The deleted measures belonged to four sub – constructs: automatic identification (knowledge of out of stock products), system (interactivity), information (understandability, currency, flexibility, and precision), and service (responsiveness, empathy, and IS training). But, as shown in the Appendix, four original constructs and their nine sub – constructs still remained distinct.

The alpha values and factor loadings at the first refinement stage were good. While high reliabilities of individual constructs are necessary conditions for the total reliability of the evaluation tool, they are not sufficient. The total reliability of the evaluation tool should be measured by a linear combination. The total reliability for four original constructs and their sub – constructs could be tested by using the formula for the reliability of linear combinations (Nunnally & Bernstein, 1994). The combined reliability values shown in the Appendix were also quite high. These high combined reliability values indicated good internal consistency among the constructs of the evaluation tool.

2) The Second Refinement Stage Using Dataset 2

The major objective of this stage is to evaluate the robustness of the evaluation tool. Cronbach' s alpha coefficient does not adequately estimate errors caused by factors external to the tool, such as differences in test situations and respondents. To assess the between – test error, a new datasetshould be used in the additional analyses (Churchill, 1979). Therefore, the evaluation tool derived the first refinement stage was re – refined with dataset 2.

Through reliability analysis and EFA using dataset 2, only accessibility measure, belonged to information sub – construct, was deleted. There were no meaningful changes in the evaluation tool derived from the first refinement

Table 5. The second refined result of the evaluation tool

Construct	Alpha value	Linear combination	Sub-construct	Alpha value	Factor loading	Linear combination
Risk	0.903	0.980	Economic risk	0.863	0.906	0.911
			Non-economic risk	0.834	0.892	
RFID	0.955		Automatic identification	0.925	0.920	0.960
			Item-level identification	0.943	0.886	
IS	0.954		System	0.894	0.952	0.955
			Information	0.921	0.773	
			Service	0.924	0.822	
Network	0.956		Stability	0.936	0.775	0.960
			Usability	0.930	0.847	

stage. Therefore, it was ascertained that the evaluation tool was amply robust. As shown in Table 5, the evaluation tool derived from the second refinement stage also had very high values of individual reliabilities and reliabilities of linear combinations. These high values of the reliabilities showed that the constructs of the second refined evaluation tool were also internally consistent.

4. Validation of the Evaluation Tool

1) The Structural Stability

To assess the stability of the derived construct structure, confirmatory factor analysis (CFA) was used. The CFA results were presented in Table 6. These results were calculated with Amos S/W (version 4.01). In CFA, an assessment of the overall model fit must firstly be done. GFI ($\geq$0.90 is recommended) and RMSEA ($\leq$0.08 recommended) were used as absolute fit measures. Some GFI values were a little low, but all RMSEA values were adequate. TLI ($\geq$0.90 recommended) and CFI ($\geq$0.90 recommended) were used as incremental fit measures, and normed χ^2($<$ 5 recommended) was used as a parsimonious fit measure. Their values were also good. Although some fit values were not good, they were acceptable. Therefore, it was concluded that the overall model fit of the evaluation tool was good.

Once the overall model fit is accepted, the measurement of each construct can then be assessed for unidimensionality and composite reliability. Unidimensionality can be tested by EFA or CFA. Unidimensionality was already ensured because the evaluation tool was refined by EFA. Therefore, only the examination of composite reliability was needed. Composite reliability was assessed by construct reliability. All values of construct reliability in Table 7 were over 0.9 and exceeded the recommended level (0.7).

The results of the overall model fit and the measurement model assessment supported the structural stability of the evaluation tool. So it was expected that the final evaluation tool shown in Table 6 was useful in many RFID - TS application areas.

2) Construct Validity

The final evaluation tool showed high reliabilities and internal consistencies.

Table 6. The Structural Stability Assessment of the Evaluation Tool by CFA (*: p 〈 0.01)

Construct and sub-construct		Measure	Construct loading	Construct reliability	Overall model fit
Risk	Economic risk	Financial risk	0.842*	0.961	GFI(0.930), RMSEA(0.061), TLI(0.955), CFI(0.967), CMINDF(1.733)
		Performance risk	0.946*		
		Time risk	0.861*		
	Non-economic risk	Psychological risk	0.761*		
		Social risk	0.713*		
		Physical risk	0.824*		
RFID	Automatic identification	Labor costs	0.827*	0.987	GFI(0.786), RMSEA(0.071), TLI(0.900), CFI(0.909), CMINDF(1.989)
		Store selling area	0.799*		
		Physical flows	0.837*		
		Shrinkage	0.737*		
		Control of the supply chain	0.869*		
		Knowledge of customer behavior	0.611*		
		Delivery disputes	0.739*		
	Item-level identification	Management of perishable items	0.791*		
		Returns management	0.877*		
		Tracing quality problems	0.888*		
		Recall management	0.823*		
		Counterfeiting	0.711*		
IS	System	Scalability	0.753*	0.990	GFI(0.770), RMSEA(0.071), TLI(0.888), CFI(0.899), CMINDF(1.992)
		Effect on job	0.632*		
		Effect on external constituencies	0.905*		
		Effect on internal processes	0.812*		
		Effect on knowledge and learning	0.793*		
	Information	Accuracy relevance	0.698*		
		Completeness	0.874*		
		Competitive intelligence	0.831*		
		Presentational quality	0.942*		
		Breadth	0.628*		
		Depth	0.672*		
	Service	Assurance	0.820*		
		Following-up	0.876*		
		Intrinsic quality of provider	0.775*		
		Interpersonal quality of provider	0.897*		
		Flexibility	0.823*		
Network	Stability	Reliability	0.943*	0.984	GFI(0.863), RMSEA(0.076), TLI(0.940), CFI(0.950), CMINDF(2.140)
		Correctness	0.936*		
		Security	0.875*		
	Usability	Response time	0.880*		
		Workload	0.884*		
		Interface	0.755*		
		Efficiency	0.874*		
		Fairness	0.607*		

However, while they are necessary conditions for the tool' s construct validity, they are not sufficient. The tool must satisfy certain other conceptual and empirical criteria to be considered as having good construct validity (Parasuraman et al., 1988).

The basic conceptual criterion pertaining to construct validity is face or content validity. Assessing content validity is qualitative rather than quantitative. It involves examining two aspects: (1) the thoroughness with which the construct to be measured and its domain are explicated and (2) the extent to which the measures represent the construct' s domain (Parasuraman et al., 1988). Churchill' s paradigm includes some methods to satisfy both requirements. And this study strictly followed Churchill' s paradigm to satisfy these requirements. Therefore, the final evaluation tool could be considered to possess content validity.

The empirical validity of the final evaluation tool was assessed by examining its convergent validity and discriminant validity. Convergent validity is assessed by t – tests for construct loadings (Anderson & Gerbing, 1988). The results of t – test were shown in Table 6 and all construct loadings were statistically significant ($p < 0.01$). These results indicated a high convergent validity of the final evaluation tool.

Discriminant validity was assessed by the average variance extracted (AVE) test. Discriminant validity is demonstrated if AVE estimates are greater than the squared correlations and exceeds 0.50 (Hair et al., 1998). The AVE test results of four constructs were shown in Table 7. All AVE estimates exceeded 0.50 and all the squares of the correlations except the square of the correlation between IS and network were smaller than their AVE estimates. The high correlation between IS and network was somewhat anticipated because it was

Table 7. Test for Discriminant Validity

Construct	AVE	Correlation			
		Risk	RFID	IS	Network
Risk	0.570	1			
RFID	0.646	0.75	1		
IS	0.667	0.55	0.69	1	
Network	0.698	0.31	0.47	0.88	1

already found in the measure derivation step that there were some conceptual redundancies between IS measures and network measures. But it could be concluded that four constructs as a whole had good discriminant validity.

All constructs of the final evaluation tool adequately satisfied the conceptual and empirical criteria (i.e., content validity, convergent validity, and discriminant validity). Summing up, these results supported that the final evaluation tool had construct validity.

5. Relative Importance of the Constructs

The primary objective of this study was to develop an evaluation tool for RFID – TS. To accomplish the objective, four constructs to evaluate RFID – TS performance were derived through Churchill' s paradigm. Four constructs were as follows: risk, RFID, IS, and network. They were divided into nine sub – constructs. The proposed evaluation tool of RFID – TS was shown as a concise multiple – item tool with good validity. Food supply chain participants can use it to better evaluate RFID – TS performance and, as a result, improve the food supply service.

To examine the relative importance of four constructs, this study regressed the overall performance measure on the factor loadings for the constructs and the sub – constructs. The regression results were shown in Table 8. The adjusted values of two regression models were statistically significant. The result of Model 1 in Table 8 came from the regression model using four constructs as independent variables. Among four constructs, IS was found to be the most influential construct in evaluating the performance of RFID – TS. The risk construct was also significantly influential. These findings were similar to those of Heo & Han (2003).

Table 8. The Relative Importance of the Constructs

Regression model	Independent variable	R^2 (Adjusted R^2)	F value	p value	Standardized beta	*t value*	Significant level	Durbin-Watson (d)	*VIF*
Model 1	Risk	0.175 (0.166)	20.829	0.000	0.214	2.868	0.005	1.861	1.332
	IS				0.267	3.579	0.000		1.332
Model 2	Economic Risk	0.177 (0.169)	21.183	0.000	0.247	3.594	0.000	1.887	1.129
	Service				0.267	3.892	0.000		1.129

Model 1 was tested to the multicollinearity and the independence of the residuals. The multicollinearity of the residuals was tested by VIF (variance inflation indicator). A common cutoff threshold of VIF value is 10.0 (Hair et al., 1998). As shown in Table 8, no VIF values exceeded 10.0. Therefore, it was demonstrated that there was no multicollinearity problem. The independence of the residuals was tested by the Durbin – Watson test which used the upper and lower critical values, D_U and D_L. If "$d > D_U$", it is concluded that the residuals are not autocorrelated. If "$d > D_L$", it is concluded that the residuals are autocorrelated. If we choose alpha = 0.05, the critical values corresponding to sample size = 200 and four constructs are $D_L = 1.758$ and $D_U = 1.779$. In this study, *d*value (= 1.861) was greater than D_U. Therefore, it was concluded that the residuals were not autocorrelated, that is, they were independent.

The results of Model 2 in Table 8 came from the regression model using nine sub – constructs as independent variables and also did not have any problems in the multicollinearity and the independence of the residuals. Economic risk (a sub – construct of risk) and service (a sub – construct of IS) were found as the significantly influential sub – constructs. The results of Model 2 were consistent with the results of Model 1. The coefficient of the service sub – construct was a little greater than the coefficient of the economic risk sub – construct. Heo & Han (2003) also found the service sub – construct as a significant performance factor. These results supported the fact that the importance of IS service was increasing (Zmud, 1984).

The consistent results of two regression analyses demonstrated that the evaluation tool was designed well. The evaluation tool for RFID – TS was designed to be applicable across a broad spectrum of food industries. So it will be useful for people in charge of evaluating RFID – TSs in most food industries.

Ⅳ. Conclusions

Recently, food safety has been a matter of primary concern among consumers. This phenomenon has made consumers desire the introduction of a traceability system (Loureiro & Umberger, 2007). If a traceability system could efficiently trace food products, it should have a capability to identify individual products.

RFID technology can give the capability to a traceability system. It is a reason why RFID – TSs are preferred to barcode – based traceability systems.

This study followed Churchill' s paradigm to develop an evaluation tool for RFID – TS. From literature reviews, the initial measures were derived. The initial measures were screened by a focus group. A total of 54 measures screened by the focus group were decreased to 45 measures through the first refinement stage. And after the second refinement stage, 44 measures remained in the evaluation tool. These measures were grouped into four constructs: risk, RFID, IS, and network.

The refined evaluation tool was reassessed by CFA on the stability of the construct structure. Then the validity test of the evaluation tool was conducted in the convergent validity and discriminant validity aspects. Most validity indices were very good. Additionally, the evaluation tool was developed through a strict construct development procedure, Churchill' s paradigm. Therefore, it seemed that the evaluation tool for RFID – TS was sufficiently valid.

Among four constructs of the developed evaluation tool, IS and risk were found as the relatively more important ones. Also, among subconstructs, service was identified as the most important subconstruct and economic risk was next. These results mean that IS service and food safety become more influential in food trace area. Most buyers are known as risk averters (Mitchell, 1999). Risk aversion is likely to be heightened in the case of food risk because the severity of the consequences are much greater than purchasing risk associated with manufactured products. Also, buyers perceive that economic related risks (money and time risks) are more important than non – economic risks (Roselius, 1971). If we take these facts consideration, we have to design RFID – TS to serve buyers flexibly and kindly and to eliminate food related risks, especially economic risks. And these findings should be reflected in developing weights of RFID – TS' s performance indices.

This study has future research topics. The evaluation tool for RFID – TS was not tested through actual RFID – TS performance datasets. The reasons for this were that RFID – TS just began to be introduced and that there were few actual RFID – TSs. Therefore, the assessment of the evaluation tool with actual datasets is remained as a future research. The development of a weighted evaluation tool is also a future research topic. This study did not consider the weights of the measures. But through regression analysis it was found that the constructs were not equally important. It implies that if the constructs are weighted differently, it is possible to evaluate RFID – TS more exactly. Finally, the development of the indices of the constructs is also an important future research topic.

Appendix: The first refined result of the evaluation tool

Construct	Alpha value	Linear combination	Sub-construct	Alpha value	Factor loading	Linear combination
Risk	0.927	0.985	Economic risk	0.873	0.840	0.929
			Non-economic risk	0.872	0.847	
RFID	0.962		Automatic identification	0.942	0.885	0.967
			Item-level identification	0.948	0.886	
IS	0.966		System	0.899	0.903	0.970
			Information	0.947	0.776	
			Service	0.926	0.738	
Network	0.96		Stability	0.93	0.785	0.967
			Usability	0.941	0.770	

References

Anderson, J. C., and D. W. Gerbing, "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach", *Psychological Bulletin*, Vol.103, No.3, 1988, pp.411-423.

Bayrak, T., and M. R. Grabowski, "Safety-critical wide area network performance evaluation," *International Journal of Information Technology & Decision Making*, Vol.2, No.4, 2003, pp.651-667.

Bellotti, V., M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. Lopes, "Making sense of sensing systems: Five questions for designers and researchers," In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, 2002, pp.415-422, Minneapolis, Minnesota, USA.

Brown, I., and J. Russell, "Radio frequency identification technology: An exploratory study on adoption in the South African retail sector," *International Journal of Information Management*, Vol.27, 2007, pp.250-265.

Chang, J. C., and W. R. King, "Measuring the performance of information systems: A functional scorecard," *Journal of Management Information Systems*, Vol.22, No.1, 2005, pp.85-115.

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

DeLone, W. H., and E. R. McLean, "Information systems success: The quest for the dependent variable," *Information Systems Research*, Vol.3, No.1, 1992, pp.60-95.

DeLone, W. H., and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *Journal of Management Information Systems*, Vol.19, No.4, 2003, pp.9-30.

DeLone, W. H., and E. R. McLean, "Measuring e-Commerce success: Applying the DeLone & McLean information systems success model," *International Journal of Electronic Commerce*, Vol.9, No.1, 2004, pp.31-47.

Flores, J. L. M., S. S. Srikant, B. Sareen, and A. Vagga, "Performance of RFID tags in near and far field," In *2005 IEEE International Conference on Personal Wireless Communications*, 2005, pp.353-357.

French, R. H., D. J. Torrieri, L. E. Miller, and J. S. Lee, "Efficient computation of network performance measures for the mobile subscriber equipment (MSE) network," In *Military Communications Conference*, Vol.2, 1990, pp.548-552.

Golan, E., B. Krissoff, F. Kuchler, L. Calvin, K. Nelson, and G. Price, "Traceability in the US food supply: Economic theory and industry studies," *Agricultural Economic Report Number* 830, 2004, Economic Research Service, US Department of

Agriculture, Washington, DC.

Haapala, H. E. S., "Operation of electronic identification of cattle in Finland," In *The Proceedings of the 4th European Conference in Precision Agriculture*, 2003, Berlin, Germany, pp. 688 - 694.

Hair, Jr., J. H., R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate data analysis*. (5th ed.). NJ: Prentice - Hall, 1998.

Heo, J., and I. Han, "Performance measure of information systems (IS) in evolving computing environments: An empirical investigation," *Information & Management*, Vol.40, 2003, pp.243 - 256.

Hou, J. L., and C. H. Huang, "Quantitative performance evaluation of RFID applications in the supply chain of the printing industry," *Industrial Management & Data Systems*, Vol.106, No.1, 2006, pp.96 - 120.

Houghton, J. R., G. Rowe, L. J. Frewer, E. Van Kleef, G. Chryssochoidis, O. Kehagia, S. Korzen - Bohr, J. Lassen, U. Pfenning, and A. Strada, "The quality of food risk management in Europe: Perspectives and priorities," *Food Policy*, Vol.33, 2008, pp.3 - 26.

Igbaria, M., and M. Tan, "The consequences of the information technology acceptance on subsequent individual performance," *Information & Management*, Vol.32, No.3, 1997, pp.113 - 121.

ISO, *ISO's strategies in detail*, 1999, http://www.iso.ch/presse/longrang.pdf

Kaplan, L. B., G. J. Szybillo, and J. Jacoby, "Components of perceived risk in product purchase: A cross validation," *Journal of Applied Psychology*, Vol.54, No.3, 1974, pp.287 - 291.

Kärkkäinen, M., "Increasing efficiency in the supply chain for short shelf life goods using RFID tagging," *International Journal of Retail & Distribution Management*, Vol.31, No.10, 2003, pp.529 - 536.

Kärkkäinen, M., and J. Holmström, "Wireless product identification: Enabler for handling efficiency, customisation and information sharing," *Supply Chain Management: An International Journal*, Vol.7, No.4, 2002, pp.242 - 252.

Koste, L. L., M. K. Malhotra, and S. Sharma, "Measuring dimensions of manufacturing flexibility," *Journal of Operations Management*, Vol.22, No.2, 2004, pp.171 - 196.

Lo Bello, L., O. Mirabella, and N. Torrisi, "Modelling and evaluating traceability systems in food manufacturing chains," In *13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2004, pp.173 - 179.

Loureiro, M. L., and W. J. Umberger, "A choice experiment model for beef: What US consumer responses tell us about relative preferences for food safety, country - of - origin labeling and traceability," *Food Policy*, Vol.32, 2007, pp.496 - 514.

Mahon, D., and C. Cowan, "Irish consumers' perception of food safety risk in minced beef," *British Food Journal*, 106, No.4, 2004, pp.301 - 312.

Miraglia, M., K. G. Berdal, C. Brera, P. Corbisier, A. Holst - Jensen, E. J. Kok, and H. J. Marvin, "Detection and traceability of genetically modified organisms in the food production chain," *Food and Chemical Toxicology*, Vol.42, No.7, 2004, pp.1157 - 1180.

Mitchell, V. W., "Consumer perceived risk: conceptualisations and models", *European Journal of Marketing*, Vol.33, No.1/2, 1999, pp.163 - 195.

Mitchell, V. W., and M. Greatorex, "Consumer risk perception in the UK wine market," *European Journal of Marketing*, Vol.22, No.9, 1988, pp.5 - 15.

Mitra, K., M. C. Reiss, and L. M. Capella, "An examination of perceived risk, information search and behavioral intentions in search, experience and credence services," *The Journal of Services Marketing*, Vol.13, No.3, 1999, pp.208 - 228.

Najjar, L. J., J. C. Thompson, and J. Ockerman, "Wearable computer for quality assurance inspectors in a food processing plant," In *Proceedings of the 1st IEEE International Symposium on Wearable Computers*, 1997, Cambridge, Massachusetts, USA, pp.163 - 164.

Nunnally, J. C., and I. A. Bernstein, *Psychometric theory*, (3rd ed.). NY: McGraw - Hill, 1994.

Parasuraman, A., V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A multiple - item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, Vol.64, No.1, 1988, pp.12 - 40.

Perrin - Martinenq, D. "The role of brand detachment on the dissolution of the relationship between the consumer and the brand," *Journal of Marketing* Management, Vol.20, No.9/10, 2004, pp.1001 - 1023.

Pitt, L. F., R. T. Watson, and C. B. Kavan, "Service quality: A measure of information systems effectiveness," *MIS Quarterly*, Vol.19, No.2, 1995, pp.173 - 185.

Rai, A., S. S. Lang, and R. B. Welker, "Assessing the validity of IS success models: An empirical test and theoretical analysis," *Information Systems Research*, Vol.13, No.1, 2002, pp.50 - 69.

Reynolds, G., and K. Lynch, *RFID: A practical approach - 7 critical success factors in RFID deployments*, 2005, http://www.sensormatic.com/

Rodrigues, M. A. "Evaluating performance of high - speed multiaccess networks," *IEEE Network Magazine*, Vol.4, No.3, 1990, pp.36 - 41.

Roselius, T. "Consumer rankings of risk reduction methods," *Journal of Marketing*, Vol.35, No.1, 1971, pp.56 - 61.

Sahin, E., Y. Dallery, and S. Gershwin, "Performance evaluation of a traceability system: An application to the radio frequency identification technology," In *2002 IEEE International Conference on Systems, Man and Cybernetics*, Vol.3, 2002, pp.653 - 658.

Scholtz, J., and S. Consolvo, "Toward a framework for evaluating ubiquitous computing applications," *Pervasive Computing,* Vol.3, No.2, 2004, pp.82 - 88.

Seino, K., S. Kuwabara, S. Mikami, Y. Takahashi, M. Yoshikawa, H. Narumi, K. Koganezaki, T. Wakabayashi, and A. Nagano, "Development of the traceability system which secures the safety of fishery products using the QR code and a digital signature," In *MTS/IEEE Techno-Ocean 2004*, Vol.1, Nov. 2004, pp.476-481.

Smith, G. C., J. D. Tatum, K. E. Belk, J. A. Scanga, T. Grandin, and J. N. Sofos, "Traceability from a US perspective," *Meat Science*, Vol.71, No.1, 2005, pp.174-193.

Taniguchi, Y., and N. Sagawa, "IC tag based traceability: System and solutions," In *Proceedings of the 21st International Conference on Data Engineering*, 2005, pp.13-17.

Van Der Spiegel, M., P. A. Luning, G. W. Ziggers, and W. M. F. Jongen, "Evaluation of performance measurement instruments on their use for food quality systems," *Critical Reviews in Food Science and Nutrition*, Vol.44, 2004, pp.501-512.

Wang, N., N. Zhang, and M. Wang, "Wireless sensors in agriculture and food industry: Recent development and future perspective," *Computers and Electronics in Agriculture*, Vol.50, 2006, pp.1-14.

Yeung, R. M., and J. Morris, "Food safety risk: Consumer perception and purchase behaviour," *British Food Journal*, Vol.103, No.3, 2001, pp.170-87.

Zhang, J., and B. Mukherjee, "A review of fault management in WDM mesh networks: Basic concepts and research challenges," *IEEE Network*, March/April 2004, pp.41-48.

Zmud, R. W., "Design alternatives for organizing information systems activities," *MIS Quarterly*, Vol.8, No.2, 1984, pp.79-93.

Developing an Evaluation Tool of RFID-based Traceability Systems

Kim, Jin-Baek

Abstract

Recently, traceability systems are introduced as a new food safety information system. To trace food products efficiently, they must have an automatic identification capability at the individual product level. This capability can be gained through RFID technology. But there is not yet any performance evaluation tool on RFID – based traceability systems (RFID – TS). This study developed an evaluation tool of RFID – TS. To develop the tool, this study considered the objective and the components of RFID – TS as their performance constructs. According to Churchill' s paradigm, the tool was established through two stages. The final evaluation tool consisted of four constructs (risk, operational benefits, IS (information system), and network) and nine sub – constructs. Among the four constructs, risk and IS were found as the most important performance constructs through regression analysis. Among the nine sub – constructs, service (which belonged to the IS construct) and economic risk (which belonged to the risk construct) were found as the most important performance sub – constructs.

key words : evaluation tool, traceability system, RFID, performance, food safety