경영정보학연구
제16권 제2호
2006년 6월

경영정보학 연구에 나타난 실증적 연구방법 적용상의 문제: 평가를 통해 얻은 교훈

강 신 철*, 이 준 기**, 최 정 일***

Application of Empirical Research Methods in Information Systems Research: Gaining Lessons Through Evaluation

Shincheol Kang, Zoonky Lee, Jeongil Choi

Application of appropriate research methods has assumed an important role in knowledge accumulation endeavors in the scientific research community. The current study reported here summarizes how we used the empirical methods in IS research, what we need to improve in using empirical study as research methodology through the set of comprehensive research guideline. From the survey and review of seminal and classical research guidelines, we developed our own 28 checklist for research design, statistical analysis, and conclusion, which can be commonly applied to all articles that employ inferential statistical methods in IS research. Then, we critically evaluated the usage of empirical research methods in major IS journals by using the checklist, with the goal of improving the quality of academic papers.

In this study, we scrutinized four major IS journals which published empirical papers from 1991 to 2000: MIS Quarterly, Journal of MIS, Information Systems Research, and Decision Sciences. As a result of intensive evaluation work, we could highlight many areas that are lagging and call for greater attention with regard to the proper usage of empirical study in IS research. The research findings in this study can be referred as checklist and guideline when IS researcher applies the empirical method.

Keywords : Empirical Study, IS Research, Research Method, Research Design Guideline

---

* 한남대학교 경영정보학과 교수
** 연세대학교 정보대학원 부교수
*** 교신저자, Merrimack College 경영학과 조교수

Ⅰ. Introduction

Application of appropriate research methods has assumed an important role in knowledge accumulation endeavors in the scientific research community. Improper application of research methods in research design, data analysis and inference often leaves us with a serious question on the credibility of their findings [Lucas, 1991]. Contributors of scientific communications are responsible for their research designs, the applicability of the statistical test used, and the validity of the conclusions drawn.

Every editorial board of journals in well-disciplined fields provide guidelines for reporting research methods [Wood, 1981; Nicholas and Katz, 1985; Bailar and Mosteller, 1988; Campion, 1993; Wilkinson and Task Force on Statistical Inference, 1999]. The field of IS research is no exception, and significant improvements in the research methods used in IS studies have been achieved in the past decade [Alavi and Carlson, 1992; Boudreau *et al.*, 2001]. The current study reported here summarizes what we have accomplished and provide future guidelines for the soundness of IS research.

The first purpose of this article is to develop the comprehensive checklist, which can be commonly applied to all articles that employ inferential statistical methods in IS research. We believe that checking the validation of research method is important as a guide for better practice, both for journal editors/reviewers and for authors. The second purpose of the study is to critically evaluate the usage of empirical research methods in major IS journals by using the checklist, with the goal of improving the quality of academic papers.

We scrutinized four major IS journals which published empirical papers from 1991 to 2000: MIS Quarterly, Journal of MIS, Information Systems Research, and Decision Sciences. Using these journals, we highlighted many areas that are lagging and call for greater attention.

Ⅱ. Previous Research

There have been numerous attempts to evaluate the methodological soundness of empirical theses in many disciplinary fields including psychology, sociology, medical science, marketing, and organizational behavior [Schor and Karten, 1966; White, 1979; Wood, 1981; LaTour and Miniard, 1983; Gaither and Glorfeld, 1985; Mitchell, 1985; Nicholas and Katz, 1985; Delucchi, 1987]. The field of IS research has also produced helpful guidelines in this area during the last decade [Mumford *et al.*, 1984; Jarvenpaa *et al.*, 1985; Baroudi and Olikowski, 1988; Straub, 1989; Pinsonneault and Kraemer, 1993; Chau, 1999; Boudreau *et al.*, 2001].

Research methodological issues can be, to a large extent, classified into three phases of research process: research design, statistical analysis, and conclusion [Lucas, 1991]. The research design phase includes decisions regarding research question, hypothesis generation, level of analysis and measurement, sampling scheme, measurement, and exploratory data analysis. The second phase, statistical analysis, entails statistical tool selection, assumption test, statistical data report, and power analysis. The final step is to draw a conclusion and discuss the result. In the last section, researchers are expected to direct future research and draw attention to limitations of the study.

Research guidance from IS studies has been centered on the design phase. Those topics discussed are atheoretical hypothesis generation [Keen, 1980; Hamilton and Ives, 1982; Jarvenpaa *et al.*, 1985; Culnan and Swanson, 1986; Lucas, 1991], inappropriate sampling, multiple level analysis, low response rate [Hughes and Gibson, 1991; Kraemer and Dutton, 1991; Pinsonneault and Kraemer, 1993], and measurement validation [Straub, 1989; Newsted *et al.*, 1991; Sethi and King, 1991; Zmud and Boynton, 1991; Boudreau *et al.*, 2001]. There are fewer discussions of other two phases. We found only a few studies which covered statistical analysis, statistical tool selection and procedure [Chin and Todd, 1995], or power analysis [Baroudi and Olikowski, 1988]. We could recognize there are few studies that discuss the final phase of drawing conclusions in a proper way.

Ⅲ. Development of Checklist for Evaluation

As shown above, efforts to improve the rigor of academic research in the field of IS are abundant, but scattered and limited in scope and range. This study expands the previous research on methodological issues to develop a comprehensive checklist of twenty-eight items. All items are based on research method guidance found in previous studies. Items were refined through several rounds of pilot studies and validation. We will explain the item development procedures in the later research method section and discuss each item in the following section, presenting previous efforts in IS research.

Phase 1: Research Design

The goal in the research design phase should be to familiarize the reader with the problem, hypothesis generation, the procedure for testing the hypothesis, and the selection of research strategy. This study generated 16 items to access the design phase and named them D1 through D16.

D1. Is the problem clearly stated?

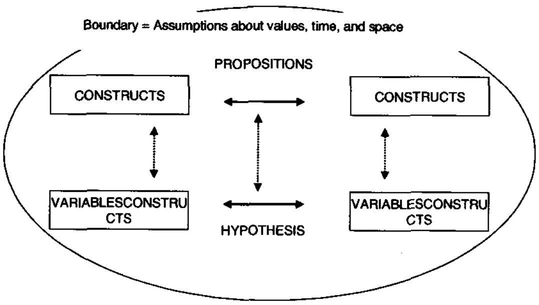
A problem is a question that asks how variables are related [Kerlinger, 1986]. Kerlinger [1986] suggested three criteria of formulating a good research problem:

(1) The problem should express a relation between two or more variables. It asks questions like: Is A related to B? How are A and B related to C?
(2) The problem should be stated in question form. Questions have the virtue of posing problems directly.
(3) It demands that the problem be such as to imply possibilities of empirical testing.

The clear statement of research quest should imply empirical testing of relationships among constructs, although it is not stated in interrogative form. Provision of a research model, typically in graphic form, can be deemed a clear problem statement.

D2. Are the research problems appropriately converted into a hypothesis?

Research questions are converted into a hypothesis, which is a conjectural statement of the relations between two or more variables. Marx [1976a] defined the hypothesis as a more tentative or provisional type of theoretical proposition, specifically, an experimental or statistical prediction in a study. Bacharach [1989] viewed a theory as a system of constructs and variables in which the constructs are related to each other by propositions and the variables are related to each other by hypothesis, as depicted in <Figure 1>.



<Figure 1> Components of a Theory

D3. Is the hypothesis stated in a relation between two or more variables?

By definition, propositions are relating constructs, not variables. Constructs may be defined as "terms which, though not observational either directly or indirectly, may be applied or even defined on the basis of the observable" [Kaplan, 1964, p.55]. A variable may be defined as "an observable entity which is capable of assuming two or more values" [Schwab, 1980]. Therefore, when a research question is transferred to a statement which intends to be empirically tested using observed data, it should be stated as a hypothesis, rather than a proposition.

D4. Does it have empirically testable hypotheses?

Hypotheses are statements of relations, and like problems, must imply the empirical testing of the stated relations [Kerlinger, 1986]. If there is no way of empirically testing a theory, it is scientifically worthless, no matter how plausible, imaginative, or innovative it may be [Marx, 1976b].

D5. Are there any multiple meanings in one hypothesis statement?

Hypothesis statements should not be ambiguous. For example, a hypothesis implicating many variables or indirect relationship in a statement like "decision performance is affected by computer literacy and/or education level" should be avoided. In studies where research problems are depicted in graphic form by a research model, the authors should formally state hypotheses, since, in many cases, a researcher does not test all theoretical relationships depicted in the diagram. Furthermore, as there is no agreement on standardized graphical representation of a research model among IS researchers, various graphical notations often confuse readers. Therefore, presenting a clear statement of hypothesis is very important to help readers understand the researchers' intent.

Pinsonneault and Kraemer [1993] reviewed 122 IS articles and found that more than half of the studies they reviewed either did not provide research hypotheses/questions or did not describe them clearly enough to get an understanding of the study's aim.

D6. Does it justify all hypotheses based on theory, logical inference, or previous research?

All hypotheses should be developed on a sound theory base. Normally, a hypothesis is generated from an existing theory or previous studies. However, mere attempts to justify hypotheses based on previous studies, without a conceptual discussion that describes why the current study meets the assumptions and definitions of a particular paradigm or theory, should be avoided [Schkade, 1990]. On the other hand, the investigator's innovative idea and hypothesis should be mentioned properly with the process of logical inference.

As for the qualified hypothesis statement, a null statement is preferred [Kerlinger, 1986]. This argument is based on the tenets of falsificationism [Rosental and Rosnow, 1991, p.35]. Falsifiability determines whether a theory is constructed such that empirical refutation is possible [Bacharach, 1989]. While the idealistic goal of science is the pursuit of universal truth, most philosophers of science would agree that theories can never be proven, only disproven [Popper, 1959]. Scientific theories can evolve as additions to, as well as replacements of, outmoded formulation. This view is supported by many IS researchers [Keen, 1980; Hamilton and Ives, 1982; Culnan and Swanson, 1986]. Even if a hypothesis is supported, the theory on which it is based is not necessarily proved to be correct because another theory might account for the results.

By rejecting null hypotheses based on the currently collected data, researchers try to accumulate knowledge of truthfulness of their proposition. Strictly speaking, scientifically valid hypotheses should be stated in null form.

D7. Is the level/unit of analysis correct?

In research on a focal unit, two types of levels exist: level of measurement and level of analysis. Level of measurement refers to the unit to which the data are directly attached, while the level of analysis is the unit to which the data are assigned for hypothesis testing and statistical analysis [Rousseau, 1987]. The level or unit of analysis can be either on the individual, system, project, group, department, company, industrial, or national level.

The level of analysis should be consistent with the level of measurement of variables included in hypotheses. The focal unit is defined as the level to which generalizations are made [Rousseau, 1987]. In practice, the focal unit often may not be identical to either the level of measurement or the level of analysis. For example, in many IS research, chief executive officers are often considered the best informants to be representative of the organization. In this case, even though the measurement and analysis are made on the individual level, generalization is made on the organizational level.

If the focal unit of research is identical to both the level of measurement and the level of analysis, evaluation is quite straightforward. Problems arise when research and theory combine different units in measurement and/or analysis [Rousseau, 1987]. Involving multiple levels of analysis often adds complexity to the sampling strategy in MIS research [Kraemer and Dutton, 1991]. When aggregating data (for example, averaging individual scores to measure department of organizational performance), special care should be exercised as to representativeness (sampling bias), inter-rater reliability, anonymity, etc.

D8. Is the linkage between target population and sampled population explicitly mentioned?

Wilkinson and a task force on statistical inference appointed by the APA Board of Scientific Affairs [Wilkinson and Task Force on Statistical Inference, 1999] stated that the interpretation of the results of any study depends on the characteristics of the population intended for analysis.

Researchers should define the population (participants, stimuli, or studies) clearly. If control or comparison groups are part of the design, which is typical in experimental designs, they should present how they are defined.

In the same vein, how a population is defined in an article affects almost every conclusion in that article. No study would intend to limit its conclusion to those specific subjects, especially in experimental studies. Therefore, researchers should always mention explicitly the linkage between the target population and the sample population. Kraemer and Dutton [1991] stated that MIS researchers need to clearly define the population to be studied.

D9. Does it use an acceptable sampling procedure?

Researchers should describe the sampling procedures and emphasize any inclusion or exclusion criteria. If the sample is stratified (e.g., by site or gender), the method and rationale for stratification ought to be properly mentioned. Pinsonneault and Kraemer [1993] found inappropriate sampling procedures in IS research. Of the articles they reviewed, almost half of the studies did not report or describe the sample frame.

In an experimental study, if random assignment is planned, the author should provide enough information to show whether or not the process is randomized. In a situation where random assignment is not feasible, the researcher needs to minimize effects of variables that affect the observed relationship between a causal variable and an outcome in sampling procedure [Kerlinger, 1986]. The author should also describe methods used to attenuate sources of bias, including plans for minimizing dropouts, noncompliance, and missing data [Rosenthal and Rosnow, 1991]. Gaither and Glorfeld [1985] examined organizational behavior journals, and found that only 30 percent of the sample used random sampling and random assignment in their sampling procedure.

D10. Does it use a sample that is appropriate for the research question and is adequately generalizable?

Appropriateness of the sample can be evaluated from two points: sample size and representativeness. First, what constitutes an appropriate sample size? The proper number of samples is determined by significance level (type I error), type of statistical tools, and number of variables in a research model [Nunnally, 1978]. The type II error also needs to be considered explicitly to determine the power of statistics and the sample size.

Tabachnick and Fidell [1989] suggested the appropriate sample size for different types of statistics. For multiple regressions, the ideal number of subjects is 20 times the number of independent variables. The minimum number of subjects required for a multiple regression is about four to five times the number of variables. For factor analyses, 200 subjects are considered to be "fair," 300 "good," 500 "very good," and 1000 "excellent." In the use of Chi-square tests, no expected cell frequency should be less than 5 and the total number of subjects should be more than 20 [Meyers and Grossen, 1974].

Generally speaking, the bigger the sample size, the better. However, due to limits of resources, researchers often doesn't suffer from insufficient sample. In the field of IS, one-half of surveys at the individual level and two-thirds at the organizational level had sample sizes of less than 150 [Pinsonneault and Kraemer, 1993].

Small sample may unduly deteriorate the quality of study results. Schor and Karten [1966] pointed out that researchers tend to place too much confidence on negative results with small-size samples. Many researchers, they argued, do not appreciate that a difference must be very great in the population before a small-sized sample will have a high probability of detecting this difference.

Secondly, characteristics of the sample may also degrade the credibility of study results, if the sample is not representative of the target population [Mitchell, 1985]. This problem often arises in experimental studies where a convenience sample is used. The suitability of students as surrogates in the decision making process, in an experimental study, depends on case-specific circumstances [Burnett and Dunne, 1988; Hughes and Gibson, 1991]. For example, if a study result is to be applied to the psychological characteristics of the general population, a student sample may be acceptable. However, if the research model is theorized in the context of business organizations, using students as surrogate subjects for the decision-maker is inappropriate. Therefore the suitability of student subjects should be clinically evaluated before testing.

D11. Does it have acceptable return rates and attrition rates, or justified properly?

In general, the higher response rates in data collection are recommended. One study suggests that response rates below 20% are undesirable [Yu and Cooper, 1983]. We realize that setting an arbitrary cutoff point without considering research method or sample method might be problematic. For example, if a sample is drawn from one or two organizations, the response rate will likely be quite high, while census type of mailing typically leads to very low rate of response. Researchers, however, should exert adequate efforts to increase return rates, addressing the influence of non-respondents and dropouts in survey design.

Pinsonneault and Kraemer [1993] found that 74 percent of the IS survey studies they examined either did not report the response rate or had a rate below 51%. Mitchell [1985] reviewed the Journal of Applied Psychology, the Academy of Management Journal, and the Organizational Behavior and Human Performance between 1979 and 1983, and found that 48 out of 126 articles failed to report response rate.

D12. Is the level of measurement appropriate for statistical analysis?

Levels of measurement, the scales associated with the levels, and the statistics appropriate to the levels are both complex and controversial problems. Kerlinger [1986] defined four general levels of measurement: nominal, ordinal, interval, and ratio.

Depending on the level of measurement, the use of certain statistics can be limited. For example, using nominal measurement, Chi-square test, run test, binomial test, or some non-parametric tests such as McNemar test and Cochran test can be used to infer some information. With ordinal scales, mostly non-parametric statistics such as Komogorov-Smirnov, Mann-Whitney, Kruskal-Wallis, Wilcoxon, or Sign test may be legitimate procedures. Strictly speaking, at least, interval or higher level ratio scales should be used for most parametric statistics, whether it is univariate or multivariate, such as correlation, regression, ANOVA, discriminant analysis, factor analysis, cluster analysis, and others. However, Kerlinger [1986, pp. 401-403] noted that "A little compromise between ordinal and interval scales is acceptable for practical purposes." In the current study, we considered the use of both Likert-type scales and the semantic differential scale as interval level of measurement.

D13. Does it justify all measures based on purpose, theory, or previous research?

If operationalization of constructs is made to collect data, the researchers are expected to summarize the psychometric properties of its scores with specific regard to the way the instrument is used in a population. Psychometric properties include measures of validity, reliability, and any other qualities affecting conclusions [Nunnally, 1978; Mitchell, 1985; Kerlinger, 1986; Rosenthal and Rosnow, 1991]. If a physical apparatus is used, the researcher should provide enough information (brand, model, design specification) to allow another researcher to replicate their measurement process.

The researcher may use standardized measures, develop measures based on literature review or based on his/her own logic. In both cases, the author should report reliability levels and include the sources of their basis. If self-developed measures are used, procedure of testing reliability and validity should be described in more detail.

D14. Does it test and report reliability and validity of measures properly?

Besides showing that an instrument is reliable, the researchers need to show that it does not correlate strongly with other key constructs. It is just as important to establish that a measure does not measure what it should not measure as it is to show that it does measure what it should. A more detailed guideline for construct validity in IS research can be found in Boudreau *et al.* [Boudreau *et al.*, 2001], Lee [Lee, 1989] and Straub [Straub, 1989].

D15. Does it have adequately detailed demographics or structure of sampled data?

To help readers understand characteristics of study subjects, the researchers are urged to report adequately detailed demographics (e.g. age, gender, status, education level, experience, workfield, etc.). Properly reported demographic data also makes replication of the study and meta-analysis possible. Again, the detailed demographic data help readers determine representativeness of samples to the target population.

D16. Does it perform an appropriate level of exploratory data analysis?

Before entering into the main statistical analysis, the investigator should conduct an appropriate level of exploratory data analysis (EDA). Listed among the most frequently used EDA techniques are univariate distribution, pair-wise scatter plots among variables, residual plots, outliers, missing data, box-plots, graphical display of multivariate profiles, etc. The researchers do not have to report all EDA results, but should mention them briefly. Gaither and Glorfeld [Gaither and Glorfeld, 1985] examined organizational behavior journals and found that only 3 percent examined the existence of outlying data points.

Phase 2: Statistical Analysis

The statistical analysis phase entails justification for selecting specific statistical tools, assumption tests, and good practices of reporting statistical information, such as p-value and power analysis. This study generated 7 items to access the analysis phase and named them A1 through A7.

A1. Are the statistical methods chosen appropriate to address the research questions or hypotheses, research design, and measure?

The appropriate statistical method depends on the questions being asked and the form of the data. Selection of appropriate statistical methods is determined by multiple factors; research problem types (difference, causality, correlation, etc.), quality of data, sample size, and the number of variables in endogenous and exogenous groups. Quality of data includes level of measurement, linearity, and distribution. For example, nominal and ordinal data require non-parametric methods, and pairing or matching should be taken into account [White, 1979]. Schor and Karten [1966] presented an anecdotal example of misuse of statistical methods:

(1) The use of chi-square when the theoretical frequency on any cell is less than five,
(2) The use of chi-square on continuous data or percentages, or
(3) The testing of differences between means rather the testing of differences against zero in a paired study.

Choosing a minimally sufficient analysis is also so important. The variety of modern statistical methods allows researchers to freely choose proper analytical tools for the research question. Although state-of-the-art methods may sometimes be necessary to address research questions effectively, simpler classical techniques often can provide sufficient answers to the research questions. Researchers should not choose analytic methods to impress readers or to deflect criticism. If the assumptions and strength of a simpler method are appropriate for data and the research problem, researchers should use that simple statistical technique rather than driving readers into the mystical labyrinth of a complex method [Wilkinson and Task Force on Statistical Inference, 1999].

A2. Is an appropriate description of the selected statistical methods provided where needed?

White [1979] suggested that in order to help non-statistician readers understand analytical procedures, researchers should provide brief descriptions of the statistical tools used. In particular, uncommonly used statistical methods, such as most nonparametric and multivariate statistical methods, require explanation for readers or reviewers to understand the purpose or procedure of applying the particular analytical tools. Even for commonly used statistical methods, explanation of the logic and application procedure of the techniques is recommended [Bailar and Mosteller, 1988].

A3. Does it address major assumptions for using a particular statistical method?

No statistical tests should be applied unless the investigator either appreciates that the assumptions underlying statistical tests are met by the data or can estimate the effect of having some assumptions unmet [Schor and Karten, 1966]. Three general assumptions about the collected data or reference population should be met in parametric statistical tests; (1) the selection of the subject from the population was random and independent, (2) the observations were drawn from the normally distributed population, and (3) the variance of each set of scores or group of scores must be comparable [Meyers and Grossen, 1974]. If one or more of the conditions are not met, the researcher should consider using nonparametric tests. If the researchers need to employ multivariate statistical techniques, in addition to linearity, normality, independence and homogeneity of observations, they should check multicollinearity and homoscedasticity [Tabachnick and Fidell, 1989; Hair *et al.*, 1998]. In particular, the use of causal inference techniques demands a very strict set of assumptions.

Data transformation can be utilized to correct violations of the statistical assumptions or to improve the relationship between variables [Hair *et al.*, 1998]. If data are manipulated or transformed, the reason and procedure should be reported.

A4. Is the particular statistical package used explicitly reported?

Kerlinger [1986] warned the researchers not to put computer programs in the foreground of research activities. There are many good computer programs for analyzing data. More important than choosing a specific statistical package is verifying the study results, understanding what they mean, and knowing how they are computed. If verifying the results by intelligent "guesstimates" is not plausible, the researcher should check them against the output of another program. Researchers should not report statistics found on a printout without understanding how they are computed or what they mean.

A5. Does the study describe resulting statistics properly? (both descriptive and inferential)

The empirical paper analyzing the collected data should appropriately report the basic statistics of data, for example, mean, standard deviation, the number of samples used in actual analysis, and the correlation/covariance matrix. These data can be utilized by other researchers for the purpose of replication or for meta-analysis purposes. Here, the author's task is to present the basic descriptive measures, point out the important comparisons, and indicate whether the differences between conditions are statistically different [Wood, 1981]. The statistical data cannot be presented in detail, and so are described by summary statistics. White [1979] suggested that a measure of the location (center) of the distribution of sample values and a measure of the dispersion (spread) be reported. A mean is an appropriate measure of location in normally distributed data, while a median is more informative in ordinal and skewed data. Presenting a correlation matrix, if available, helps readers understand the structure of observed data and enables other researchers to plunge it into further analysis.

All data needed to lead or justify the conclusions should be given [White, 1979]. Omission of relevant data diminishes readers' confidence on the study results. However, presenting overly precise data (for example, $p < 0.00001$, as is output by computer package) may unduly impress the correctness of the results or cause the reader to place greater importance on the results [O'Quigley and Baudoin, 1985].

A6. Is the effect size reported and discussed (power analysis)?

Researchers who try to induce conclusions from empirical data using statistical tools are always expected to present the effect sizes for primary outcomes. Reporting effect sizes enables readers to evaluate the stability of results across samples, designs, and analyses. It also informs power analyses and meta-analyses needed in future research. Interval estimates should be given for any effect sizes involving principal outcomes. Intervals for correlations and other coefficients of association or variation should be provided wherever possible [Rosenthal and Rosnow, 1991]. It is worth noting Cohen's anecdotal statement regarding power analysis [Cohen, 1988]:

> "Document the effect sizes, sampling and measurement assumptions, as well as analytic procedures used in power calculations. Because power computations are most meaningful when done before data are collected and examined, it is important to show how effect-size estimates have been derived from previous research and theory in order to dispel suspicions that they might have been taken from data used in the study or, even worse, constructed to justify a particular sample size. Once the study is analyzed, confidence intervals replace the calculated power in describing results".

A7. Does it properly report significance level (or p-value) and confidence interval as needed?

There are two basic philosophies underlying statistical testing using p-value. First, under Neyman-Pearson's testing hypothesis theory, the author's interest is to make decisions by testing hypotheses; therefore, the author must set a minimum thrust level beforehand. Then, if the data does not meet his/her thrust level, the author fails to reject null hypotheses. On the other hand, under the philosophy of Fisher's test of significance, the author is not interested in making his/her decision based on the p-value. So he/she report the p-value as is and leave the discretion of judgment to the reader so that the reader can reject the null hypothesis or accept alternate hypotheses based on his/her own threshold (alpha level). In the same vein, Bailar and Mosteller [1988] suggested that exact p-values, rather than statement like "p<0.05" or "p not significant," be reported where possible so that readers can compare the calculated value of P with their own choice of critical values.

Phase 3: Draw Conclusion

Campion [1993] provides three sections of the discussion and conclusion: explanation of results, derivation of implications and description of limitations. This study generated 5 items to access the conclusion phase and named them C1 through C5.

C1. Does it report the results in an appropriate manner (table, figure, graph)?

To increase readability, authors are recommended to use description aids like graphics, summary tables, or appropriate diagrams [Bailar and Mosteller, 1988]. The results have to be organized in such a way that the reader is not confronted with a large mass of data. The task is to give a simple, clear, and reasonably complete account of the results. Improper use of tables or charts may mislead readers in interpreting study results [Schor and Karten, 1966].

If there are many numbers to present (say, ten or more), a table or figure probably should be used [Wood, 1981]. The author should not present the same measure in two different ways in the article as this may confuse the reader.

C2. Is the conclusion drawn consistent with the statistical results?

The author should not over-interpret or under-interpret data and results. She/He should check the correctness of directions (positive or negative), effect-size, p-value, multi-group comparisons, etc. The author should explicitly separate objective result descriptions from subjective interpretation. The author should not overlook or minimize findings contrary to hypotheses.

The meaning of a significance level (or p-value) was often misinterpreted by some authors. The p-value does not give the probability that there is no difference, but rather the probability of the observed outcome, if the null hypothesis is true [White, 1979]. Improper conclusions may be drawn in spite of the application of proper statistical tests [Schor and Karten, 1966]. A good example is a study in which the investigator, using the proper test, calculated a probability value which turned out to be between 0.01 and 0.05 and then stated that he was 99% confident that the hypothesis was true. In this case, proper interpretation should go as follows: If the hypothesis tested were true, then the probability of obtaining sample results as bad as or worse than those observed is between 0.01 and 0.05.

Statistical significance should not be overly emphasized in the IS research field. The significance of knowledge accumulation in IS field is difficult to evaluate. In the current study, due to its subjective nature, the level of contribution of the article to the IS field was not assessed. However, excessive self-appraisal of small effects that are statistically significant should be avoided [Wood, 1981].

C3. Does it try to link findings back to original hypotheses and purposes of the study?

For the integrity of writing of academic paper, conclusions are to be in line with original research problems and hypotheses. It is not a good practice to simply express such as "accept the null hypothesis" [Wilkinson and Task Force on Statistical Inference, 1999].

C4. Does it provide a realistic delineation of limitations and weaknesses?

The author should honestly delineate limitations of the study in order to alert readers not to over or under-interpret study results. Relating research findings to those of other studies is also strongly recommended [Wilkinson and Task Force on Statistical Inference, 1999].

C5. Does it provide logical and innovative directions for future research?

From the Popperian falsification perspective, a theory is never completed with one study result. Almost no studies ever finish the stream of research, but only add one piece of knowledge to the previous repository of knowledge in that paradigm. Therefore, researchers are expected to provide new logical directions for future research [Wilkinson and Task Force on Statistical Inference, 1999].

Ⅳ. Research Method

Journal/Article Selection

Considering the period of inquiry for the current research, which covers the 10 years from

1991 to 2000, this study selected journals representative of the research conducted during this period. As a guide in selecting these journals, we used Straub [1989], Nord and Nord [1995], and Boudreau *et al.* [2001]'s studies. We selected three journals among Nord and Nord's 1st tier list and whose acceptance rate is below 20% and primary reader type is academic: MIS Quarterly, Decision Sciences, and Journal of MIS. Then we added, as Boudreau *et al.* did, a fourth journal, Information Systems Research, which was founded in 1991 and generally recognized as comparable to MIS Quarterly in quality.

Articles from these four journals were reviewed and read by three authors for a period of inquiry starting from in January 1991 and ending in December 2000. The qualifying criteria for the sample were that the article (a) tested the relationships among constructs and (b) made some inference from empirically collected data, and (c) used inferential statistical methods. These criteria are equivalent to the "theoretically grounded positivist" in Orlikowski and Baroudi's [1991] classification of research epistemology. Therefore, the major research approaches of our samples are either field studies or experimental studies. We included articles that did not explicitly set hypotheses but tried to empirically test the theoretical relationships among variables. However, we did not include case studies because even though they attempt to draw some conclusion from empirical data, they neither set hypotheses nor use inferential statistics due to their exploratory nature and lack of intent to generalize. We also excluded instrument development and descriptive studies, as they are neither intended to test the relationships among theoretical variables nor to draw conclusions from empirical data.

Item Development and Coding Procedure

Initially, we developed over thirty check items based on previous studies and evaluated sample articles based on the following scales:

NG: Neglected (coded 1)
PE: Performed Erroneously (coded 2)
PU: Performed Unsatisfactorily (coded 3)
PC: Performed Correctly (coded 4)
NA: Not Applicable (coded 5)

Then we followed several pilot studies and refinement procedures to agree on the meaning of items and evaluations among three researchers. First, we selected four articles from all four journals and evaluated those articles based on the items by three researchers. A few disagreements were found in evaluating items regarding sampling procedure, generalizability, response rate, exploratory data analysis, and significance level among three researchers. Through discussion, several ambiguous expressions were reworded and items that demanded too much subjective judgment were omitted. Second, we repeated the process with different articles to come up with agreement in evaluating each item. Again, a few disagreements on evaluations were resolved through this process. As a result, we developed written criteria for each item. For instance, if pre-testing procedure is mentioned, we evaluated Item D9 as performed satisfactorily, if not, performed unsatisfactorily. If the context of the research model implies a business environment, an undergraduate sample was rated as "Perform Unsatisfactorily" on Item 10. Some disagreements on Items D9, D10, and D11 were resolved by differentiating experimental studies from field survey studies. To our knowledge, most experimental studies reported virtually no subject dropouts during the procedure. Therefore, we decided not to evaluate Item D11 for all experimental studies, which we labeled "Not Applicable." As to Item D16, either a data screening procedure or a brief scan of a correlation matrix is decided to be satisfactory. With regard to Item A7, whether the writer choose either Fisher's or Neyman-Pearson's perspective, if F or t-statistics had been reported, even though row p-value was not given, we agreed to rate the item "Perform Satisfactorily." More detailed criteria for evaluations are shown in Appendix A. Finally, we developed the twenty-eight item checklist and conducted one final round of pilot testing. This time the three researchers agreed on items over 97% for four journals, and we finalized the items and criteria for evaluations.

In addition to the checklist, we coded the type of research (field survey or experimental), statistical methods, and unit of analysis. These three attributes are also used to evaluate the articles based on the checklist. We did not differentiate field experiments (quasi-experiment) from laboratory experimental studies, because this attribute is only used to evaluate sampling procedures and representativeness of samples. To enrich our study, we surveyed which statistical procedures are commonly used to test hypotheses in inferential studies. Unit of analysis is a great concern in validating the quality of research. Therefore, the analysis unit of our samples can be on a personal level, a group or team level, or an organization level. The unit of analysis can also be a project or an information system.

Inter-rater Reliability

In order to measure inter-rater reliability, the three-rater-cross-check-technique was employed. By alternating odd and even years for each journal, three researchers evaluated about half of the journals together and the rest of the articles for one researcher only (i.e. odd years of MISQ and JMIS and even years of ISR and DS are evaluated by all three researchers). Inter-rater reliability was tested based on those 149 articles evaluated by all three researchers. First, the average raw percentage of agreement for each item was 86%. The average agreement for each section was 85%, 85% and 88% for research design, statistical analysis and draw conclusion section, respectively. Those numbers are very high compared with other studies such as Boudreau *et al.* [2001], considering that we are testing all three researchers agreements. Along with raw agreement measures, we also tested the reliability using Cohen's [1960] Kappa coefficient. The average kappa was 0.72, which was above the generally acceptable range of 0.7 recommended by Gardner [1995] and Landis and Koch [1977]. We also calculated z-value of kappa to test the significance of the coefficients based on Siegel and Castellan [1988, p. 289]. The average z-value was 10.2 and the z-values for all 28 items were significant at the α=0.01 significance level, demonstrating that the raters exhibit significant agreement on their ratings. <Table 1> shows how the relevant studies choose and sort articles based on the research purpose and how they assure inter-rater reliability using Cohen's Kappa coefficient. Since we demonstrated reasonable range of the inter-rater reliability, one researcher reconciled the disagreements and we further performed the analysis.

<Table 1> Use of Inter-rater Reliability

Research	Main Content / Data Sorting	Inter-rater Reliability
Alavi and Carlson [1992]	◦ To classify MIS research strategies ◦ Make master list to decide which article should be included	◦ Cohen's Kappa: Not used
Boudreau *et al.* [2001]	◦ 193 article used for analysis ◦ Validate IS research via content/construct validity, reliability, and manipulation validity ◦ No description about the pre-test process to check the difference of rating among raters	◦ Percentage of agreement for the eleven coded attributes ◦ Cohen's Kappa: 0.75 (Average) ◦ Used Criteria: > 0.7 [Miles and Huberman 1994; Landis and Koch 1977; Bower and Courtright, 1984]
Moore and Benbasat [1991]	◦ Level of agreement in categorizing items for developing instrument to measure the perce-ption of IT adoption ◦ Four sorting round; use four or five external rater in each round	◦ Inter-judge raw agreement score ◦ Cohen's Kappa: 0.70 < K < 0.92 ◦ Used Criteria: > 6.5 [Vessey 1984; Jarvenpaa 1989; Todd and Benbasat, 1989]
Vessey *et al.* [2002]	◦ Pre-test: Recognize that 60 papers out of 100 papers show difference in rating.	◦ Percentage of agreement ◦ Cohen's Kappa: 0.72 < K < 0.82 (including z-value and p-value)

V. Data Analysis

A total of 335 articles are evaluated in the study. Among them 146 articles are experimental studies and the rest of 189 are field studies. 157 articles used individuals as their unit of analysis, and the rest of articles used either groups or organizations as their unit of analysis. The following table shows the number of articles by each journal and by year.

Since the main purpose of this article is to identify areas that require more attention from IS research, we choose 7 items (25% of 28 items) that are evaluated worst (see <Table 3>). A complete list of items and their evaluation is in Appendix A. For better statistical analysis of the data, we calculated the mean value of each item coding based on neglected as 1 and performed correctly as 4, excluding NA (not applicable).

<Table 2> Number of Articles per Year per Journal

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	Total	%
MISQ	3	8	4	7	4	12	7	12	8	9	74	22
ISR	4	10	10	10	11	9	5	8	9	10	86	26
JMIS	11	16	11	9	8	21	12	10	9	10	117	35
DS	2	8	2	7	4	5	12	4	13	1	58	17
Total	20	42	27	33	27	47	36	34	39	30	335	

<Table 3> Critical Issues in IS Research Method

(Unit: Number, (%))

Items	1(Ng)	2(Pe)	3(Pu)	4(Pc)	Mean (Std. Dev)
D8. Is the linkage between target population and sampled population explicitly mentioned?	188 (56%)	1 (0%)	27 (8%)	116 (35%)	2.21 (1.42)
D14. Does it test and report reliability and validity of measures properly?	89 (27%)	0 (0%)	22 (7%)	201 (60%)	3.07 (1.34)
D16. Does it perform an appropriate level of exploratory data analysis?	186 (56%)	0 (0%)	26 (8%)	121 (36%)	2.25 (1.43)
A2. Is an appropriate description of the selected statistical methods provided where needed?	158 (47%)	0 (0%)	33 (10%)	143 (43%)	2.48 (1.43)
A3. Does it address major assumptions for using a particular statistical method?	227 (68%)	0 (0%)	10 (3%)	96 (29%)	1.92 (1.37)
A4. Is the particular statistical package used explicitly reported?	291 (87%)	0 (0%)	5 (1%)	36 (11%)	1.36 (0.96)
A6. Is the effect size reported and discussed (power analysis)?	262 (78%)	0 (0%)	6 (2%)	33 (10%)	1.37 (0.97)

The average mean of those 7 items was 2.09, which is sharply contrasted with the average of 3.22 for all items. Next, we were interested in whether there were mean differences across different journals. One-way factor analysis shows that there are mean differences (with α=0.05) for 7 items among 28 items. Interestingly enough, only one of the above seven items (item D14) is significant, indicating that the above list contains issues commonly found across all IS journals that are subject to study.

Since we have evaluated items across a period of 10 years, we are interested in whether the problem issue list has changed during that period. We divided the period into the first 5 years and the last 5 years and conducted analysis again. Based on the average below 3 mean criteria, we chose the issue list for the first and later 5 years. The issue list for the first 5 years is D8, D14, D16, A2, A3, A4, A6 and C4, which is exactly consistent with the above list plus one more item, C4: (Does it provide a realistic delineation of limitations and weaknesses?). The issue list for the last 5 years contains D8, D16, A2, A3, A4 and A6. Therefore, two items, Item D14, which asks whether reliability/validity of measures reported and Item C4, which asks whether there was a realistic delineation of limitations and weaknesses, are excluded from the later list. All other items are common in both the first and last 5 years of list. <Table 4> shows the mean change on those items from the first 5 years and the last 5 years.

<Table 4> Mean Changes in the Critical Issues

Items	D8	D14*	D16*	A2	A3	A4*	A6*	C4*
F-5yrs	2.32	2.92	2.53	2.63	2.01	1.23	1.23	2.94
L-5yrs	2.12	3.21	2.01	2.36	1.86	1.46	1.49	3.23

*: Significant difference at α=0.05

The above table shows that the scores of all the issue items have stayed unchanged or improved statistically except D16 (asking whether appropriate exploratory analysis is reported), which was worsened. Particularly, along with Items D14 and C4 explained above, Items A4 and A6 (asking if the statistical package used was reported and whether the effect size was reported, respectively) improved significantly. However, in general, those items are evaluated very low in both periods and our original list of issue items stayed the same.

To better understand the nature of the problems regarding the items, we conducted further analyses with those items by type of research (i.e. experimental vs. field studies). The following table shows the mean average of each item only when the mean of item is below three.

<Table 5> Critical Issues by Research Design

(Unit: Mean Average)

Items	D8*	D14*	D15*	D16*	A2*	A3	A4	A6*
Experimental Study	1.66	2.82	2.74	1.92	2.16	1.85	1.27	1.23
Field-Study	2.63	3.25	3.39	2.49	2.73	1.98	1.42	1.49

*: Significant difference at α=0.05

The table again shows that the identified above items plus just one more item, D15 (demographics or structure of sampled data reported). The analysis shows that experimental studies have more problems regarding those items, as their mean average is significantly lower than the average scores from field-studies.

Ⅵ. Discussion

This study generated a total twenty-eight items of checklist to evaluate research methods used in four major IS journals during the last ten years. The checklist contains not only the statistical methods but also ways to generate hypotheses, research design and draw conclusions from empirical studies. Through the analysis we identified seven items that are continuously evaluated low, therefore requiring more attention from IS research.

Issue 1 : Linkage between target population and sampled population

We found that more than 65% of published papers either failed to explicitly report the relationship between the target population and sampled population or unsatisfactorily reported the relationship. Since we took the conservative evaluating approach that once the relationship is explicitly mentioned even in a sentence, it was regarded as reported, we suggest that the problem might be more serious than our scores indicated. In most cases, contrary to recommendations from other studies [Kraemer and Dutton, 1991; Pinsonneault and Kraemer, 1993; Wilkinson and Task Force on Statistical Inference, 1999], the repopulations of the study were not clearly mentioned, and even when it was mentioned, the relationship of sampled data to the population was not clearly demonstrated. This makes the external validity of the studies dubious in the conclusion.

For many field survey studies, convenience samples such as individual data from one accessible organization were used without defining the populations. We found that the problem is more serious with experimental studies. Many experimental studies have used students as their samples, but failed to explicitly mention what the population of the studies was.

While it is recommended that the probability sampling by which the sample is randomly selected and the subjects have the same probability of being selected, should be required, several forms of non-probability sampling such as convenience sample, quota sample, and purposive sample are often unavoidable [Kerlinger, 1986].

In some situations, a researcher needs to survey one or a limited number of organizations rather than all of the applicable organizations due to theoretical purposes or location limitation. A researcher may use convenience samples for data availability. The researcher may also opt for a "purposive sample" (sometimes called a theoretical sample) in which firms are selected because they exhibit features (e.g. high or low computerization and so forth) of central concern to the researcher.

Using a convenience sample does not automatically disqualify a study from publication. However researchers using non-probability samples have to understand biases resulting from such research designs and have to minimize potential problems. At minimum, the researcher should be sure to make that procedure clear to the readers.

There are some techniques to mitigate problems related to non-probability samples. Sometimes by explicit comparison of sample characteristics with those of a defined population across a wide range of variables, the case for the representativeness of a convenience sample can be strengthened [Wilkinson and Task Force on Statistical Inference, 1999].

In conclusion, if a study needs to use a convenience sample on a theoretical basis or due to other research limitations, it is desirable to clearly mention why the convenience sample is necessary and how the convenience sample represents the target population.

Issue 2 : Reporting reliability/validity of measures

There has been a plethora of research indicating that IS empirical researchers need to pay more attention to validating measurement scales [Straub, 1989; Kraemer and Dutton, 1991; Newsted *et al.*, 1991; Zmud and Boynton, 1991; Chau, 1999; Boudreau *et al.*, 2001]. Straub [1989] surveyed 117 MIS articles published in MIS Quarterly, Communications of the ACM, and Information and Management covering the period from 1985 to 1988, and found 62 percent of the studies lacked even a single form of instrument validation, then asked more attention from MIS researchers. They found that only 17 percent of the studies assessed reliability, which is the most frequent validation method, and only 14 percent mentioned construct validity. Consequently they suggested that researchers (1) pretest and pilot test instruments, and (2) journal editors encourage or require researchers to prepare an "Instrument Validation" subsection of the Methodology section. Chau [1999] reviewed 418 reliability coefficients from 63 articles published in four major IS journals between 1983 to 1995, and found that reported coefficients ranged from 0.27 to 0.99 with a mean of 0.81, which is above the frequently referenced guideline [Nunnally, 1978]. Newsted *et al.* [1991] reviewed 35 journals in which IS topics appeared from 1977 to 1988, and found only 29 percent (199 out of 672) reported reliability, 27 percent reported validity, and only 16 percent reported pretest results.

Thanks to those efforts we found that the problem of not reporting reliability/validity has been improved significantly during the last ten years.

In fact, analyzing the data from the last 5 years alone tell us that it is not a major problem as their scores have improved from 2.92 to 3.21. The percentage of studies that have not reported any basic instrument validation decreased from 32% to 22%. Although there is a sign that the problem has been alleviated during the last 10 years as it has improved from 2.6 to 3.03 (it is significant with F=3.07 with α=0.1) for experimental studies, we believe that there still is room for improvement. We, especially find that some experimental studies using theory of planned behaviors [Ajzen, 1985, 1991] or technology acceptance model [Davis, 1989] have set good exemplars that very rigorous reliability and validity tests have been reported using confirmatory factor analysis [Adams *et al.*, 1992; Agarwal and Prasad, 1998].

Issue 3 : Appropriate level of exploratory data analysis (EDA) reported

The primary goal of doing EDA is twofold; data screening and economic use of statistical method [Tabachnick and Fidell, 1989]. Careful analysis of data leads to better prediction and more accurate assessment of dimensionality [Hair *et al.*, 1998]. EDA can prevent researchers from mistakenly adopting improper analytical tools. For instance, an approximate U-shaped relation between stress level and work performance may be revealed easily from pair-wise scatter plot, but a simple regression (which requires a linear relationship between variables) may end up with very low R-square. Another reason for recommending EDA is that a lot of information can be found in simple statistics, which nullify delving into complex statistical methods.

As was suggested by Tabachnick and Fidell [1989], our evaluation on the item is based on whether a simple statistic of EDA is reported. Only correlations matrix or simple mean and standard deviations of measurements used is typically reported in the paper. Nevertheless, we find that 56% of studies have no mention of EDA, and the problem has worsened during the last ten years as their mean changed from 2.53 to 2.01. In fact, it is the only item that the scores have significantly decreased when we compared the first and last 5 years. And the problem was serious for both experimental and field studies although it was a bit more serious on experimental studies (average is 1.92 and 2.49, respectively). Along with this problem, we also find that many experimental studies failed to report adequately detailed demographics of sampled data (D15).

Issue 4 : Appropriate description of the selected statistical methods provided where needed

Studies have argued that appropriate description of the selected statistical methods should be reported even for simple statistics [Bailar and Mosteller, 1988]. We found that 87% of the article ignored the description of statistics they use. Since the journals we investigated are targeting academicians who are equipped with basic statistical method skills, the high rate of ignorance might not account for the seriousness of the problem. But we strongly believe that studies that used non-common statistical methods such as non-parametric or multivariate analysis need to provide proper description on the statistical methods they use.

Issue 5 : Addressing major assumptions for using a particular statistical method

Researchers should take appropriate level of efforts to assure that the underlying assumptions required for the analysis are logically given the data [Schor and Karten, 1966]. Gaither and Glorfeld [1985] examined 1,102 organizational behavior articles, and found that only 17 percent of the sample satisfactorily stated the assumptions of the parametric statistical test and justified its use, and only 8 percent formally tested the assumptions. Similarly our results show that 68% of studies totally ignore the assumptions, and only less than 30% of studies have correctly reported assumptions for statistical methods they used. We adopted conservative evaluations that when simple statistics like regression or t-tests are used, we marked as "performed correctly". But we find that the percentage of studies that failed to address the issue of assumptions regarding the statistical methods increased from 64% to 71% when we compare the first and last 5 years, suggesting no sign of improvement. We also find that the problem is equally serious for both experimental and field studies, as their average evaluation score is 1.85 and 1.98, respectively.

Issue 6 : Particular statistical package used explicitly reported

In order to help readers replicate and verify the statistical analytical procedure, reporting exact version and name of particular statistical packages is important [Bailar and Mosteller, 1988]. We believe that the report of statistical package is more important for studies where some specific types of statistical packages are used. But we find that 87% of studies have not reported it. Especially, we suggest that current popular use of structural equation models such as PLS, EQL or AMOS warrants report of their use of statistical packages and versions if possible.

Issue 7 : the effect size reported and discussed (power analysis)

Since the seminal article by Baroudi and Orlikowski [1988], there has been a growing concern in the IS community about the importance of power analysis. Our results reflect the trend as the evaluation scores increased from 1.23 to 1.49, which is significant at the 95% level. Despite the increase, the percentage of studies of not reporting at all in the last 5 years is still high on 73% although it is an improvement from 83.8% in the first 5 years.

Ⅶ. Limitation and Further Research

Although the current study provide IS researchers with guidance and pointed out important issues in the current practices of IS research, our results must be viewed in light of the study's limitations. First, although we could identify areas that are relatively lagging, some objective guidelines of coding prevent researchers from further identifying more issues. For instance, the item like "Is problem clearly stated (listed D1)" and "Is the conclusion drawn consistent with the statistical results? (listed C2)" are likely to subjective judgments, and we adjusted the evaluation based on objective criteria and often times the criteria allowed us to judge whether it is done or not rather than how it is done. While it was necessary to ensure the objective evaluation, future studies can address the issues more carefully by designing various aspects of evaluation criteria. Second, even though we sampled from leading IS journals, the generalizability of the results is limited to only these journals. Further research is needed to determine the applicability of these results to other journals.

Ⅷ. Conclusion

This study developed the checklist for IS research and applied it to the last ten years of IS research. We found some significant improvements in research methods. Especially, we found that more studies have begun to report construct validity for measures and statistical power during the last ten years. We also identified some problematic areas in research design. Particularly, we found that IS studies are not clear in establishing the link between their samples used in the study and population. We also recommend IS studies to report more on the assumptions of statistical methods used in the study and appropriate level of exploratory data analysis.

〈References〉

[1] Adams, D.A., Nelson, R.R., and Todd, P.A., "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication," *MIS Quarterly*, Vol. 16, No. 2, 1992, pp. 227-247.

[2] Agarwal, R. and Prasad, J., "The Antecedents and Consequents of User Perceptions in Information Technology Adoption," *Decision Support Systems*, Vol. 22, No. 2, 1998, pp. 15-29.

[3] Ajzen, I., "From Intensions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Eds.)," *Action Control: From Cognition to Behavior* (pp. 11-39). New York: Springer, 1985.

[4] Ajzen, I., "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, 1991, pp. 179-211.

[5] Alavi, M. and Carlson, P., "A Review of MIS Research and Disciplinary Development," *Journal of Management Information Systems*, Vol. 8, No. 4, 1992, pp. 45-62.

[6] Bacharach, S.B., "Organizational Theories: Some Criteria for Evaluation," *Academy of Management Review*, Vol. 14, No. 4, 1989, pp. 496-515.

[7] Bailar, J.C., III., and Mosteller, F., "Guidelines for Statistical Reporting in Articles for Medical Journals," *Annals of Internal Medicine*, Vol. 108, 1988, pp. 266-273.

[8] Baroudi, J.J. and Olikowski, W.J., "The Problems of Statistical Power in MIS Research," *MIS Quarterly*, Vol. 13, No. 1, 1988, pp. 88-106.

[9] Boudreau, M., Gefen, D., and Straub, D. W., "Validation in IS Research: A State-of-the Art Assessment," *MIS Quarterly*, Vol. 25, No. 1, 2001, pp. 1-16.

[10] Bowers, J.W. and Courtright, J.A., *Communication Research Methods*, Glenview, IL, Scott, Foresman, 1984.

[11] Burnett, J.J. and Dunne, P.M., "An Appraisal of the Use of Student Subjects in Marketing Research," *Journal of Business Research*, Vol. 14, No. 4, 1988, pp. 329-343.

[12] Campion, M.A., "Article Review Checklist: A Criterion Checklist for Reviewing Research Articles in Applied Psychology," *Personnel Psychology*, Vol. 46, 1993, pp. 705-718.

[13] Chau, P.Y.K., "On the Use of Construct Reliability in MIS Research: A Meta-Anal-

ysis," *Information and Management*, Vol. 35, No. 4, 1999, pp. 217-227.

[14] Chin, W.W. and Todd, P.A., "On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS research: A Note of Caution," *MIS Quarterly*, Vol. 19, No. 2, 1995, pp. 237-246.

[15] Cohen, J., "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, Vol. 20, 1960, pp. 37-46.

[16] Cohen, J., *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Erlbaum, 1988.

[17] Culnan, M. and Swanson, E., "Research in Management Information Systems, 1980-1984; Points of Work and Reference," *MIS Quarterly*, Vol. 7, No. 3, 1986, pp. 286-303.

[18] Davis, F.D., "Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology," *MIS Quarterly*, Vol. 13, No. 3, 1989, pp. 319-339.

[19] Delucchi, K.L., "The Use and Misuse of Chi-Square: Lewis and Burke Revisited," *Psychological Bulletin*, Vol. 94, No. 1, pp. 1987, 166-176.

[20] Gaither, N. and Glorfeld, L., "An Evaluation of the Use of Tests of Significance in Organizational Behavior Research," *Academy of Management Review*, Vol. 10, No. 4, 1985, pp. 787-793.

[21] Gardner, W., *On the Reliability of Sequential Data: Measurement, Meaning and Correction*, Mahwah, NJ: Erlbaum, 1995.

[22] Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C., *Multivariate Data Analysis* (5th ed.), Upper Saddle River, NJ: Prentice Hall, 1998.

[23] Hamilton, S. and Ives, B., "MIS Research Strategies," *Information and Management*, Vol. 5, No. 4, 1982, pp. 339-347.

[24] Hughes, C.T. and Gibson, M.L., "Students as Surrogates for Managers in a Decision-making Environment: An Experimental Study," *Journal of Management Information Systems*, Vol. 8, No. 2, 1991, pp. 153-166.

[25] Jarvenpaa, S., "The Effect of Task Demands and Graphical Format on Information Processing Strategies," *Management Science*, Vol. 35, No. 3, 1989, pp. 285-303.

[26] Jarvenpaa, S., Dickson, G., and DeSanctis, G., "Methodological Issues in Experimental IS Research: Experiences and Recommendations," *MIS Quarterly*, Vol. 9, No. 2, 1985, pp. 237-246.

[27] Keen, P., *MIS Research: Reference Disciplines and a Cumulative Tradition*, Paper presented at the The First International Conference on Information Systems, Philadelphia, PA, 1980.

[28] Kerlinger, F.N., *Foundations of Behavioral Research* (3rd ed.), New York: CBS College Publishing, 1986.

[29] Kraemer, K.L. and Dutton, W.H., "Survey Research in the Study of Management Information Systems," In K. L. Kraemer (Ed.), *The Information Systems Research Challenges: Survey Research Methods* (Vol. 3, pp. 3-57), Cambridge, MA: Harvard Business School Press, 1991.

[30] Landis, J. and Koch, G.G., "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, Vol. 33, 1977, pp. 159-174.

[31] LaTour, S.A. and Miniard, P.W., "The Misuse of Repeated Measures Analysis in

Marketing Research," *Journal of Marketing Research*, Vol. 20, No. 1, 1983, pp. 45-47.

[32] Lee, A.S., "A Scientific Methodology for MIS Case Studies," *MIS Quarterly*, Vol. 13, No. 1, 1989, pp. 33-49.

[33] Lucas, H.C., "Methodological Issues in Information Systems Survey Research," In K. L. Kraemer (Ed.), *The Information Systems Research Challenges: Survey Research Methods* (Vol. 3, pp. 273-282). Cambridge, MA: Harvard Business School Press, 1991.

[34] Marx, M.H., "Formal Theory," In M. H. Marx and F. E. Goodson (Eds.), *Theories in Contemporary Psychology* (2nd ed., pp. 234-260), New York: Macmillan, 1976a.

[35] Marx, M.H., "Theorizing," In M. H. Marx and F. E. Goodson (Eds.), *Theories in Contemporary Psychology* (2nd ed., pp. 261-286), New York: Macmillan, 1976b.

[36] Meyers, L.S. and Grossen, N.E., *Behavioral Research: Theory, Procedure, and Design*, San Francisco, CA: W.H. Freeman and Company, 1974.

[37] Miles, M.B. and Huberman, A.M., *Qualitative Data Analysis: An Expanded Sourcebook*, Thousand Oaks, CA: Sage Publications, Inc, 1994.

[38] Mitchell, T.R., "An Evaluation of the Validity of Correlational Research Conducted in Organizations," *Academy of Management Review*, Vol. 10, No. 2, 1985, pp. 192-205.

[39] Moore, G.C. and Benbasat, I., "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research*, Vol. 2, No. 3, 1991, pp. 192-222.

[40] Mumford, E., Hirschheim, R., Fitzgerald, G., and Wood-Harper, A.T. (Eds.), *Research Methods in Information Systems*. Amsterdam, The Netherlands: North-Holland, 1984.

[41] Newsted, P.R., Munro, M.C., and Huff, S.L., "Data Acquisition Instruments in Management Information Systems," In K. L. Kraemer (Ed.), *The Information Systems Research Challenges: Survey Research Methods* (Vol. 3, pp. 187-209), Cambridge, MA: Harvard Business School Press, 1991.

[42] Nicholas, J.M. and Katz, M., "Research Methods and Reporting Practices in Organizational Development: A Review and Some Guidelines," *Academy of Management Review*, Vol. 10, No. 4, 1985, pp. 737-749.

[43] Nord, J.H. and Nord, G.D., "MIS Research: Journal Status Assessment and Analysis," *Information and Management*, Vol. 29, No. 1, 1995, pp. 29-42.

[44] Nunnally, J.C., *Psychometric theory* (2nd ed.), New York: McGraw-Hill, 1978.

[45] O'Quigley, J. and Baudoin, C.E., "Null Hypothesis and the Misuse of Statistics," *Nature*, 316, 1985.

[46] Orlikowski, W.J. and Baroudi, J.J., "Studying Information Technology in Organizations: Research Approaches and Assumptions," *Information Systems Research*, Vol. 2, No. 1, 1991, pp. 1-28.

[47] Pinsonneault, A. and Kraemer, K.L., "Survey Research Methodology in Management Information Systems: An Assessment," *Journal of Management Information Systems*, Vol. 10, No. 2, 1993, pp. 75-105.

[48] Popper, K., *The Logic of Scientific Discovery*. New York: Harper and Row, 1959.

[49] Rosental, R. and Rosnow, R.L., *Essentials of Behavioral Research: Methods and Data Analysis* (2nd ed.), McGraw-Hill, 1991.

[50] Rousseau, D.M., "Issues of Level in Organizational Research: Multi-level and Cross-level Perspectives," *Research in Organizational Behavior*, Vol. 7, 1987, pp. 1-37.

[51] Schkade, D.A., "Prospects for Experiments Focusing on Individuals in IS Research," In I. Benbasat (Ed.), *The Information Systems Research Challenges: Experimental Research Methods* (Vol. 2, pp. 49-52), Cambridge, MA: Harvard Business School Press, 1990.

[52] Schor, S. and Karten, I., "Statistical Evaluation of Medical Journal Manuscripts," *Journal of the American Medical Association*, Vol. 195, No. 13, 1966, pp. 1123-1128.

[53] Schwab, D.P., "Construct Validity in Organizational Behavior," In B. W. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 2, pp. 3-43), Greenwich, CT: JAI Press, 1980.

[54] Sethi, V. and King, W., "Construct Measurement in Information Systems Research: An Illustration in Strategic Systems," *Decision Sciences*, Vol. 22, No. 3, 1991, pp. 455-472.

[55] Straub, D.W., "Validating Instruments in MIS Research," *MIS Quarterly*, Vol. 13, No. 2, 1989, pp. 147-167.

[56] Tabachnick, B.G. and Fidell, L.S., *Using Multivariate Statistics* (2nd ed.), New York: Harper & Row, 1989.

[57] Todd, P.A. and Benbasat, I., "A Critique of Cognitive Styles Theory and Research," In E. R. McLean (Ed.), *Proceedings of First International Conference on Information Systems* (pp. 82-90), Philadelphia, PA, 1980.

[58] Vessey, I., "An Investigation of the Psychological Processes Underlying the Debugging of Computer Program," Unpublished PhD Dissertation, Department of Commerce, University of Queensland, 1984.

[59] Vessey, I., Ramesh, V., and Glass, R.L, "Research in Information Systems: An Empirical Study of Diversity in the Discipline and Its Journal," *Journal of Management Information Systems*, Vol. 19, No. 2, 2002, pp. 129-174.

[60] White, S.J., "Statistical Errors in Papers in the British Journal of Psychiatry," *British Journal of Psychiatry*, Vol. 135, 1979, pp. 336-342.

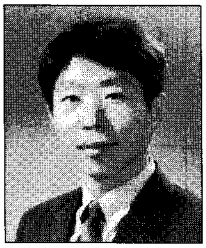
[61] Wilkinson, L. and Task Force on Statistical Inference, A. B. o. S. A., "Statistical Methods in Psychology Journals: Guidelines and Explanations," *American Psychologist*, Vol. 54, No. 8, 1999, pp. 594-604.

[62] Wood, G., *Fundamentals of Psychological Research* (3rd ed.), Boston: Little-Brown, 1981.

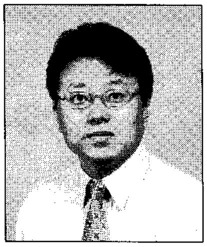
[63] Yu, J., and Cooper, H., "A Qualitative Review of Research Design Effects on Response Rates to Questionnaires," *Journal of Marketing Research*, Vol. 36, 1983, pp. 36-44.

[64] Zmud, R.W. and Boynton, A.C., "Survey Measures and Instruments in MIS: Inventory and Appraisal," In K. L. Kraemer (Ed.), *The Information Systems Research Challenges: Survey Research Methods* (Vol. 3, pp. 149-180), Cambridge, MA: Harvard Business School Press, 1991.

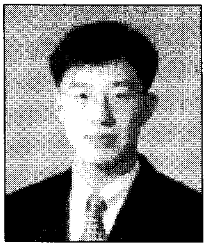
◆ 저자소개 ◆



강신철 (Kang, Shincheol)
현재 한남대학교 경영정보학과 교수로 재직 중이며, 고려대 경영학과를 졸업하고, 뉴욕주립대에서 경영정보학 석사, 네브라스카주립대에서 경영정보학 박사학위를 취득하였다. 경력으로는 한양투자금융(현 하나은행)에서 4년간 재직하였고 목원대학교에서 조교수로 5년간 근무하였다. 주요 관심분야는 정보기술이 조직에 미치는 영향, 업무 프로세스혁신, 중소기업의 정보화 혁신, 지식경영 시스템의 구축방안 등이며 이와 관련한 실무 컨설팅과 연구를 진행하고 있다.



이준기 (Lee, Zoonky)
현재 연세대학교에서 정보대학원 부교수로 재직 중이며, 서울대 계산통계학과를 졸업한 후, 미시간대학에서 통계학 석사, 카네기멜론대학에서 의사결정론 석사, 그리고 남가주 대학에서 경영정보학 박사학위를 취득하였다. 경력으로는 네브라스카 주립대학에서 조교수로 4년을 재직하였고 쿠퍼스앤드라이브랜드 경영컨설턴트로 3년을 근무하였다. 주요 관심분야는 정보기술을 통한 조직의 전략 재구성이며 특히 새로운 B2B 모델, 가격정책 모델, 채널 모델, 전략수립 모델에 관심을 가지고 연구를 진행하고 있다.



최정일 (Choi, Jeongil)
현재 미국 Merrimack College 경영학과 조교수로 재직 중이며 충남대학교 경영학과를 졸업한 후, 서울대학교에서 경영학 석사 그리고 University of Nebraska-Lincoln에서 경영정보학 박사학위를 취득하였다. 경력으로는 정보통신정책연구원(KISDI) 그리고 프랑스 인시아드(INSEAD)에서 연구원으로 근무하였다. 주요 관심분야는 전자상거래 모델연구 및 경영전략, 정보통신기술의 응용, 정보시스템 이용의 성과관리 등이다.

◆ 이 논문은 2005년 1월 24일 접수하여 1차 수정을 거쳐 2006년 6월 2일 게재확정되었습니다.