Communications of the Association for Information Systems

Volume 16 Article 45

December 2005

External Validity in IS Survey Research

William R. King
University of Pittsburgh, billking@katz.pitt.edu

Jun He
University of Pittsburgh, JunHe@katz.pitt.edu

Communications of the Association for Information Systems

EXTERNAL VALIDITY IN IS SURVEY RESEARCH

William R. King
Jun He
Katz Graduate School of Business
University of Pittsburgh
bill.king@katz.pitt.edu

ABSTRACT

This study focuses on the issues of external validity, coverage error and nonresponse error in IS survey research. Data from the empirical papers in 5 years of issues from three "A rated" IS journals are presented and analyzed. Recommendations are made based on the analyses, basic scientific principles and the authors' experience and judgment.

Keywords: survey research, research methods, external validity, coverage error, nonresponse error

I. INTRODUCTION

External validity is an important determinant of the usefulness of survey research results. Coverage error and nonresponse error importantly influence external validity. This criticality is not always recognized in IS research, even that which is published in the highest-rated journals.

The purpose of this paper is to:

1. review the underlying concepts of external validity, coverage error, and nonresponse error and
2. identify the frequency with which coverage and nonresponse errors are dealt with and reported in IS empirical research.

The analytic methods that are used to assess nonresponse error and the level of detail given to the discussion of the issue of these potential errors in study results are also treated. The primary objective is to make recommendations based on these results, on generally-accepted scientific criteria, and on our judgment in order to promote maximizing external validity in IS survey research.

These assessments are not straightforward; Campbell and Stanley [1996, p. 17] indicated that "... the problems of external validity are not logically solvable in any neat, conclusive way". Groves and Lyberg [1988] state that, "There are so many ways of calculating, response rates that

comparisons across surveys are fraught with complications" [p. 28]. This latter comment was made in the context of telephone surveys, but it is generally considered to be valid across a wide range of survey contexts.

We adopted the definition of survey research used by Pinsonneault and Kraemer [1993] which focuses on the purpose of the survey being to produce quantitative descriptions of some aspect of the populations studied by asking people structured and predefined questions using a sample of the population. Thus, we limit ourselves to these issues as they relate to positivist research [Lee and Baskerville, 2003].

We expect that our survey-related analyses and recommendations apply to the contexts of in-person surveys (such as student surveys), mail, e-mail and telephone surveys of pre-identified parties, surveys of parties identified only by their position in an organization, and personal survey interviews, although we do not discuss each of these specific contexts in detail.

II. BACKGROUND

Survey research is a major presence in Information Systems (IS). In a recent review of the use of different methodologies, Plavia and colleagues (2004) examined the articles published at seven "leading" IS journals during the years 1993-2003, and found that about 22% articles employed survey methods to gather data; in contrast, only 11.6% articles employed Frameworks and Conceptual Models - the second-most-widely-used methodology in their analysis.

The significant role of survey research may reflect the increased attention that MIS researchers place on the generalizability of their research findings. MIS is an applied discipline, and relevant research with generalizable results is widely advocated for its potential impacts to the business world [Zmud, 1996; Davenport and Markus, 1999].

> *"Because the field of information systems (IS) is not just a science but also a profession (and therefore has professional constituents such as IS executives, managers, and consultants), the generalizability of an IS theory to different settings is important not only for purposes of basic research, but also for purposes of managing and solving problems that corporations and other organizations experience in society" [Lee and Baskerville, 2003; p. 221].*

Compared with other methodologies with controlled settings, such as laboratory and field experiments, survey research involves examining a phenomenon in a wide variety of natural settings [Pinsonneault and Kraemer, 1993]. Therefore, this method has the potential to produce generalizable results that can be applied to populations other than the sample tested. However, such potential needs to be carefully assessed and "designed into" the study.

Generalizability is a broad concept that can be addressed from various perspectives [Lee and Baskerville, 2003]. From a coverage perspective, the generalizability of survey research is most discussed as the extent to which the tested sample represents a population of interest by presenting a similar set of key characteristics. It is to this generalizability issue, referred to as external validity - and two sources of error[1], coverage error and nonresponse error, that this study is addressed.

[1] Dillman [2000] identified four sources of error common in a survey research: sampling error, measurement error, coverage error, and nonresponse error. Sampling error can occur when a sample is drawn from a population. Large samples mitigate, but do not eliminate, its impact. Measurement error is induced by the makeup of the instrument used to collect data. It can be reduced through careful pre-testing and pilot testing to enhance clarity and reduce ambiguity and it can be assessed and enhanced through statistical techniques. These two sources of error are considered irrelevant to the coverage perspective and therefore not addressed in this study.

EXTERNAL VALIDITY

External validity refers to the generalizability of sample results to the population of interest, across different measures, persons, settings, or times. External validity is important to demonstrate that research results are applicable in natural settings, as contrasted with classroom, laboratory, or survey-response settings.

With respect to external validity, the ideal research procedure involves selecting a random sample from the relevant real-world[2] population, operationalizing research variables to parallel those in the real-world, selecting a research setting that is representative of the environmental variation in the real-world, and choosing a design that preserves the correspondence between the two settings and which provides the type of information required for decision-making [Calder et al., 1981].

In a typical IS research study, external validity is often explicitly or implicitly sacrificed for "... the greater statistical power that comes through having isolated settings, standardized procedures, and homogeneous respondent populations" [Cook and Campbell, 1979]. Nonetheless, if such a tradeoff is made, good scientific practice suggests that it should be as explicit as possible and that the tradeoff itself should be evaluated through such stratagems as analyzing potential coverage error and nonresponse error.

COVERAGE ERROR

Coverage error is introduced when the frame from which the sample is drawn does not include all of the relevant characteristics in the population to which inferences are to be drawn [Dillman, 2000]. Coverage error affects the external validity of the results of studies

- that rely on self-reporting by individual respondents, whether the measures being reported are factual (e.g. age) or perceptual (e.g. responses on a Likert scale), and
- in which the population to which inferences are to be drawn may be individuals operating in the "natural" world (e.g. managers or IT professionals) or organizations for whom individual responses are taken to, or aggregated to, be representative.

Coverage error contributes to the questioning of external validity. For example, sample frames use students

- who are also managers or professionals (e.g. part-time MBA students),
- who previously held such positions (e.g. many full-time MBA students) or
- who never held such positions (e.g., most undergraduates),

in making estimates or explanations that are taken to reflect a population of full-time managers, professionals, or real-world organizations.

The survey method may introduce coverage error, as in the use of on-line surveys in which individuals who do not have access to, or regularly use, e-mail or the WWW are ignored and those who do are over-sampled [Wade and Parent, 2001]. Sometimes, this coverage error becomes apparent when respondents are first contacted by non-electronic means and then given the option to respond by either electronic or non-electronic means. For instance, King et al. [2002] reported that more than half of the presumably-computer-literate chief knowledge officers who were given this option chose to respond in hard-copy form.

[2] We use the term real world to refer to what goes on outside academia, other research settings, or video games and other entertainment.

The basic method to address coverage error is to ensure that there is sufficient correspondence between relevant aspects of the population and the sample frame to expect the effects observed in the sample to be observed in the total population.

Dealing with coverage error in the context of results that are to be applied to practicing managers, professionals, or organizations can be achieved by

- less reliance on convenience samples or
- ensuring greater correspondence by such approaches as by sampling part-time MBAs who are practicing managers rather than full-time MBAs who are not, or
- redefining of populations (e.g., as management students, rather than managers), or "correspondence procedures" [Calder et al., 1981].

Correspondence procedures may be a replication of the study with different subgroups of the target population [e.g., Palmer, 2002], or purposive sampling of individuals that characterize the population and who are known to differ on important dimensions [e.g., Susarla et al., 2003] or by sampling only the most prevalent type of individual in the population [Cook and Campbell, 1979].

The importance of coverage error is amplified by the heavy reliance on some form of convenience sampling in IS survey research. Or course, convenience sampling is used in other methodological contexts as well. For example, in experimental research, the coincidence of a given set of students taking a class that happens to be offered in a particular term. In case study research, an opportunity may arise to observe directly and to participate in a phenomenon being studied [e.g., Bechky, 2003] or to perform interviews in which subjects are asked to recount past experiences [e.g., Orlikowski, 2002].

In IS research, a common convenience sampling situation occurs when management students or IT students are presumed to represent practicing managers or IT professionals. Some studies use full-time managers or professionals who are enrolled in executive education programs on the assumption that they are closer to the population of all managers or all IT professionals than are full-time students [e.g., Earl, 2001]. Similarly, students enrolled in masters-level programs or executive programs may be thought to be more like full-time professionals than are undergraduate students.

A second-level of coverage error is introduced when individual respondents, or some subgroup or sample of them, are taken to represent an organization in which they participate. This potential difficulty is often-unremarked-upon in many studies. Because organizations, by definition, cannot respond to any stimulus, it is often assumed that the perceptions of an individual (e.g. the CIO) or the aggregated perceptions of a set of individuals (e.g.: "users") reflect the organization's response. In such instances, sometimes the consistency of multiple respondents from the same organization is tested rather than just aggregated into an overall response [e.g., Armstrong and Sambamurthy, 1999].

Coverage error may be further amplified when a focal individual (e.g. a CIO) is asked to identify others to be respondents (e.g. users). Even when specific directions are given about how the selection is to be made, the likelihood that a non-representative sample is chosen is increased [e.g., Chang and King, 2005].

The general problem of convenience sampling may be best understood if one begins with the premise that he/she wishes to test hypotheses or theories related to persons, systems or organizations using students, or other readily available surrogates as respondents. While the idea that a study would begin with this premise is at odds with some basic conventions of scientific inquiry, it is nonetheless the case that the confluence of an idea, theory, or testable hypothesis and the opportunity to conduct a convenience sample of students or managers participating in an educational program is often the starting point for IS research studies.

More generally, with this starting point, the researcher may identify a series of background variables, or constructs. Such factors are not included in a theory and/or may not be intended to

be measured in a study. However, if included, the factors might interact with the theoretical variables and influence the results significantly.

An infinity of such background factors occur in any study. Campbell and Stanley [1966], for example, say that an internally-valid design applies only to (experimental groups) of "... a specific age intelligence, socioeconomic status, geographic region, historical moment, orientation of the stars, or orientation of the magnetic field, barometric pressure, gamma radiation, etc," [p. 17]. However much this statement may trivialize such factors, their possible significance, and therefore the possible impact of excluding them, is amplified in the convenience sampling situation. It is impossible to include, or to assess the effect of excluding, all such background factors. Yet, in convenience sampling, the difference between the units of analysis in the sample frame and the units in the population to which inferences are to be drawn may be of critical importance.

Even when the students used as respondents are practicing managers or professionals who are enrolled in a seminar or short course, it is reasonable to surmise that the fact that they enrolled means that they are less-than-representative of all such individuals. For example they may be more inquisitive, or more open to change

This way of reasoning about coverage error is similar to issues involving control variables. Control variables may be considered to be those:

- that may not be of direct importance to the theory and/or to the focal relationships being tested, and/or
- that are background variables, such as environmental or demographic factors, and/or
- that may be analyzed only in a manner that complements the main analysis (e.g., entered into a regression model that supplements the primary regression model), and/or
- whose relationship to the dependent variable may have been established empirically previously.

Thus, control variables in survey research are tested to see:

- if they mimic control variables in experimental research, or
- if they verify the impact of previously-identified explanatory variables that are not of primary significance to the study in question, or
- if they reduce error variance, or
- if they moderate effects in structural equation modeling [e.g., Nidumolu and Subramani, 2003].

Therefore, in IS survey research, control variables often are given an intermediate position between pure background variables (which most often are not measured) and those which are included in the theory that is utilized and/or are incorporated into hypotheses. Thus, analyzing background variables as control variables in survey studies represents one way of dealing with some of the coverage issues that may be important.

NONRESPONSE ERROR

While response rates do not directly affect the validity of the statistical analysis of survey data, low response rates are often associated with small sample sizes. small sample sizes negatively influence statistical power, may preclude the use of certain statistical techniques (e.g, LISREL), and increase the size of confidence intervals. Low response rates also influence the risk of nonresponse-induced errors. Perhaps more importantly, low response rates may seriously influence the perceived credibility of studies' results [Luong and Rogelberg, 1998].

Nonresponse error is introduced if non-respondents are different from respondents in terms of characteristics that are relevant to the study [Dillman, 2000]. In general, error due to nonresponses is presumed to be in direct relationship to increases in the rate of nonresponse and the level of variation in the true attitudes, beliefs or perceptions that are being assessed [American Association for Public Opinion Research, 2004].

Nonreponse error can call the external validity of a study into question. Empirical evidence shows that well-accepted methods of assessing nonresponse error such as early/late comparisons or follow-ups with nonrespondents may be valid and reliable [Linder et al., 2001]. However, the widespread perception is that, in IS, as in some other areas of social science, those methods are not widely followed and/or discussed in research papers [Richardson, 2000], thereby contributing, at least, to a loss of credibility. This study seeks to assess the validity of those perceptions (Section III).

Nonrespondents may be classified as active and passive. Active (purposeful) non-respondents may decide that completing a survey is too time-consuming, or irrelevant to their job or organization or just that "I get too many surveys." The authors encountered situations in which surveyed individuals say that their company policy is not to respond to external surveys. Passive nonrespondents intend to respond, but forget or "just didn't get to it."

The authors found that their own survey response rates are directly associated with the perceived quality and relevance of the research instrument, particularly, it's relevance to the potential respondent's job or current issues. In some of the studies that we did [e.g., Chang and King, 2005], numerous respondents requested permission to reproduce and distribute the survey instrument in their organization. When told that the instrument was not "final" (because the measures needed to be assessed statistically for validity and reliability), most requested that the final version be sent as soon as possible. In such instances, an instrument that is perceived to be useful by the respondent, if only because it reminds them of the multiple dimensions of the issue that is the focus of the survey, can serve as a basis for further data gathering in an organization.

The four basic ways to assess nonrespondent error are:

archival, follow-up,

wave, and intentions [Rogelberg et al., 2003].

In the archival approach, comparisons are made between the respondents and the population in an archival set. such comparisons are often not feasible, but it can be done in the context of ongoing research programs [Gannon et al., 1971].

In the follow-up approach, a small sample of nonrespondents is surveyed, often by telephone. They are then compared with respondents. The problem with this approach is that a second level of nonrespondents is usually created, and the researcher is left to wonder whether these double-non-responders created significant biases [Sosdian and Sharp, 1980].

In the wave approach, early and late responders are compared on the assumption that late responders are more like nonrespondents than are early responders. This approach further assumes that late responders are "reformed" passive nonrespondents [Ellis et al., 1970]. Alternately, the number of days to respond may be included in a regression equation.

The intentions approach to assessing error is related to the follow-up approach in that it involves comparisons of the attitudes of those not intending to respond with those who do intend to response [Rogelberg et al., 2003]. In effect, this approach constitutes a two-stage survey. Some studies showed differences in personality traits, such as approval-seeking and authoritarianism, between these groups [e.g., Rosenthal and Rosnow, 1969]. This approach is more feasible within an organization than in a more general context. In multi-organizational survey contexts, the closest approximation is instances in which nonrespondents are asked to give reasons for their failure to respond (See Table 6 in section III).

III. METHODOLGY

The methodology used here is adapted from that used by Smith [2002] in his study of nonresponse error reporting in leading social science journals. It reflects the notion that "Every academic field is marked by its literature" [McLean, 1996; p. 151].

Three journals, generally considered to be "A-level" in the IS field, were selected for assessment ─ *Management Information Systems Quarterly (*MISQ*), Information Systems Research (*ISR), and the *Journal of Management Information Systems* (JMIS) [Vessey et al., 2002][3]. Articles from these three leading IS journals are considered to reflect some common research practices in our field. We believe that this pragmatic approach is defensible even though some top IS papers appear in journals that are not solely IS journals and that some excellent IS papers appear in IS journals that are not top-rated by all. We expect that studies appearing in these three top journals reflect the enacted best research practices of the field.

We selected the time period Spring, 1999 through Spring, 2004[4] somewhat arbitrarily, because we wished to make the assessment on a recent/current and not historical, basis. All 21 issues of each of the 3 journals for this time period were inspected. The 199 empirical articles of the total of 411 that were published in that period were reviewed carefully.

We developed a coding scheme to indicate how external validity, coverage error, and nonresponse error were treated in each of these empirical papers. Part of the coding scheme is that used by Smith [2002], except that he was concerned with whether papers noted their adherence to AAPOR [2000] standards, something that is unlikely to be noted in IS papers. His scheme was therefore initially partially adopted and then complemented with other categories that appeared relevant as the review progressed.

IV. RESULTS

Data Collection Method. Table 1 shows the data collection methods used in the 199 empirical studies. Online, in-class and mail surveys comprised slightly in excess of half of the empirical papers. The *MIS Quarterly* had the highest proportion of survey papers and the lowest proportion of laboratory experiments. We found this to be somewhat surprising since MISQ originated, and remains affiliated with, the University of Minnesota, whose MIS researchers were once well-known for lab experimentation [e.g., Gallupe et al., 1988; Watson et al., 1988; Zigurs et al., 1988].

Table 1. Research Methods

Data Collection Methods	Number of Empirical Studies (Percentage)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
Survey	27 (58.7%)	27 (44.3%)	44 (47.8%)
Field Experiment	2 (4.3%)	1 (1.6%)	3 (3.3%)
Laboratory Experiment	12 (26.1%)	19 (31.1%)	27 (29.3%)
Secondary Data	5 (10.9%)	14 (23%)	14 (15.2%)
Interview	0 (0%)	0 (0%)	4 (4.3%)
Total	46	61	92

[3] Other "A journals," such as *Management Science* (MS) and *Organizational Science* (OS), were initially intended to be included in the study, but were subsequently excluded because they publish papers on a wide range of topics beyond IS and because the judgmental assessment of which articles qualified to be IS articles was so subjective as to be indefensible. This decision was at some odds with the method used by Vessey et al [2002] who "... coded only those articles (from MS and *Decision Sciences*) that we considered to be IS articles." [p. 136] Our different judgment is, probably in part, due to the rapid expansion of the scope of the IS field since their 1995-99 assessment. We found that we could not, a priori, develop a set of selection guidelines for *MS* and *OS* that was satisfactory when applied.

[4] *MISQ,* (23)1-(28)1; ISR*,* (10)1-15(1); JMIS (15)4-(20)4.

Most surveys were conducted by mail. On-line and in-class surveys were somewhat significant in number in ISR (4 and 5 respectively) but not so much in MISQ (1 and 4 respectively) and JMIS (2 and 1 respectively).

Level of Study. Table 2 classifies the 199 papers according to the level of the study. Most studies were performed at the individual level. Organizational-level studies ranked second. MISQ publishes more than three times as many individual-level studies as organizational-level ones, whereas ISR and JMIS focus. relatively more on organizations

Table 2. Study Level of Analysis

Study Level	Columns Under Each Journal Show No. of Empirical Studies (%)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
National	0 (0%)	0 (0%)	1 (1.1%)
Industry	0 (0%)	0 (0%)	0 (0%)
Organizational	9 (19.6%)	15 (24.6%)	25 (27.2%)
Group	5 (10.9%)	7 (11.5%)	20 (21.7%)
Individual	31 (67.4%)	39 (63.9%)	43 (46.7%)
Literature Review	1 (2.2%)	0 (0%)	3 (3.3%)
Total	46	61	92

Coverage Error. Table 3 addresses the treatment of coverage error. Even though coverage error is usually an important determinant of the generalizability of the finding, over 40% of papers in all three journals ignore coverage error and almost 20% only mention it briefly. Comparatively, Malhotra and Grover [1998] found that 80% of the 25 survey papers from the top four Production and Operations Management (POM) journals "defined and justified the sample frame."

Table 3. Coverage Error

Treatment of Coverage Error	Number of Empirical Studies (Percentage)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
Discuss Coverage Error	20 (43.5%)	20 (32.8%)	37 (40.2%)
Mentioned, but Not Discussed	8 (17.4%)	12 (19.7%)	15 (16.3%)
Ignoring, or Not Discussing, Coverage Error	18 (39.1%)	29 (47.5%)	40 (43.5%)
Total	46	61	92

Of the articles that discussed coverage error, the discussions were mostly presented in their research limitation sections. Statements like "the sample of the study only includes … (e.g., geographic area, Fortune 1000, industry type, or certain organizational/group/individual context) and thus may limit the generalizability of the results" were not rare. We found the positively-tuned justification of the sample frame, such as Thong's [1999] "comparing the profile of this study's sample with other available data on IS adoption...This figure corresponded closely with the national average...giving further confidence about the representativeness of the sample" (p.199-200), an exception.

In addition, these articles addressed coverage error with various approaches, e.g., demographic analysis, statistical testing using control variables, descriptively explaining sampling procedures/rationales, to name a few. No commonly accepted procedures and methods exist for conducting a coverage error analysis. Therefore, we could not asses the practice further.

Table 4 addresses response rate reporting for the 81 mail surveys in the study set. Non-mail surveys were omitted from this table because 100% response rates can be achieved in-class and because response rates may not be meaningful in on-line surveys. All three journals had more than 80% of studies that both defined how response rate was calculated and gave numerical values.

Table 4. Overview of Mail Survey Response Information

Treatment of Response Rates	Number of Mail Survey Studies (Percentage)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
No Information	1 (4.5%)	0 (0%)	5 (12.2%)
Partial Information	3 (13.6%)	2 (11.1%)	3 (7.3%)
Defined and Reported	18 (81.8%)	16 (88.9%)	33 (80.5%)
Total	22	18	41

Table 5 shows the reported numerical values of response rates for mail surveys. The average response rates reported are generally in keeping with “rules of thumb” that are well known in survey circles [e.g., Yu and Cooper, 1983], although each journal published some papers with low response rates (as low as approx. 8%).

Table 5. Mail Survey Statistics

Survey Data	Mail Survey Responses Data (Standard Deviations)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
Average Number of Surveys Mailed	1119 (785)	710 (572)	1015 (1325)
Average Number of Surveys Received	743 (562)	376 (397)	432 (493)
Average Usable Sample Size	290 (203)	280 (304)	261 (329)
Lowest Response Rate Reported	11.0%	15.1%	7.8%
Highest Response Rate Reported	84.0%	89.0%	85.5%
Response Rate (mean, median, and standard deviation)	35.7%, 26.1% (24.4%)	44.6%, 43% (23.0%)	33.3%, 26% (23.3%)

Note: Numbers in parenthesis indicate standard deviation.

Table 6 shows the reported treatment of nonresponse error analysis. The top four categories are taken from Dooley and Lindner [Dooley and Lindner, 2003]. The other categories were added by this paper’s authors to provide richness to the description. All four basic methods (wave, follow-up, archival, and intentions) were used. However, the somewhat-startling result shown in Table 6 is that well over half of all studies in all three

Table 6. Nonresponse Analysis Methods

Nonresponse Analysis Method	Number of Mail Survey Studies (Percentage)		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
Comparison of Early to Late Respondents (Wave)	4 (18.2%)	2 (11.1%)	5 (12.2%)
Using "Days to Respond" as a Regression Variable (Wave)	0 (0%)	0 (0%)	2 (4.9%)
Compare Respondents to Nonrespondents (Follow-up)	3 (13.6%)	1 (5.6%)	7 (17.1%)
Compare Respondents on Characteristics Known a Priori (Archival)	2 (9.1%)	2 (11.1%)	3 (7.3%)
Assess Reasons for Non-response (Intentions)	2 (9.1%)	1 (5.6%)	1 (2.4%)
No Mention Nor Analysis	8 (36.4%)	10 (55.6%)	22 (53.7%)
Mentioned, but Not Analyzed	2 (9.1%)	2 (11.2%)	1 (2.4%)
Other	1 (4.5%) *	0 (0%)	0 (0%)
Total	22	18	41

Note: * used data from a previous study

journals (approx. 48%- 70%) neither mentioned nor performed nonresponse analysis. This outcome appears to reflect a much-higher disregard for nonresponse error than is reported in some other social science fields [e.g., Dooley and Lindner, 2003; Malhotra and Grover, 1998].

Table 7 shows that a high proportion of survey studies reported on the distribution of respondents by industry, personal demographics or firm demographics (revenue, number of employees, etc). Some of them reported more than one of these sets of data, although most of these studies did not analyze these data for assessment purposes.

Table 7. Analyses of the Sample

Analyses Conducted	Number of Mail Survey Studies (Percentage)*		
	MIS Quarterly	Information Systems Research	Journal of Management Information Systems
Industry pattern analysis	7 (25.9%)	7 (25.9%)	14 (31.8%)
Demographic analysis	20 (74.1%)	17 (63%)	28 (63.6%)
Firm Size, Number of Employees	6 (22.2%)	5 (18.5%)	15 (34.1%)

Note: * Columns in this table do not sum to same values as in previous tables because some studies performed multiple analyses and some reported none.

V. CONCLUSIONS AND RECOMMENDATIONS

That we reviewed articles published by only three leading IS journals during Spring 1999 - Spring 2004 no doubt introduced coverage error to this study. Given the respected ratings of the three journals, we are confident that their publications reflect “the best research practices” in IS research. However, the reader should not generalize the results to other IS journals without careful examination of their articles.

In drawing conclusions and making recommendations, we adopt the point of view that a journal’s values, standards, and publication criteria are reflected in the papers that it publishes. Thus, we make both statements of fact (e.g., “ISR publishes 5 papers per year”) and statements that reflect this perspective (e.g., “MISQ appears to maintain higher standards for...”). We believe that while this paper gives general guidance to all reviewers and editors about these matters, the best way to assess the standards of each journal is by assessing the published portfolio.

The results in several of the tables (1, 2, and 7) reflect distributions of the foci of research and survey respondents in the three journals. However, the other tables portray a rather-surprising disregard for external validity, as reflected in the treatment of coverage error and nonresponse error.

To ensure the external validity, or generalizability, of findings from a survey research setting to a natural setting, every researcher should address the question of whether the results would have been the same if a 100% sample of the population had been conducted or if repeated random samples were drawn. Unfortunately, in a very high proportion of survey research studies published in three high-rated journals in the IS field, the external validity of the findings is unknown.

Even in those studies that addressed nonresponse error, only a few explicitly questioned the external validity of the results. Generally, no differences were found in comparing respondents and nonrespondents using any of the four most common (archival, follow-up, wave, and intentions) approaches. This outcome may imply that studies in which external validity is shown to be compromised are not submitted to, or are rejected by, top journals. However, the large proportion of published studies that ignored the issue suggests that this conjecture may not be correct. A more positive hypothesis is that nonresponse error was checked, but not reported in some papers.

Of course, studies that ignore external validity might do so because their authors believe that it is unimportant or because they are aware that their study fails in this regard. Both of these reasons are unacceptable

Based on the results of this study, our review of the relevant literature in other fields, and considerable experience-based judgment, we offer the following recommendations:

1. That, insofar as possible, acceptance criteria for the publication of survey-research based papers in IS journals should clearly require explicit treatment of the external validity of findings; in particular as reflected by analysis and discussion of coverage error and nonresponse error.
2. When it is treated, coverage error is usually discussed in a Limitations section. The tone of coverage discussions in almost all of the papers assessed was negative or defensive. Instead, researchers should strive to establish the best-possible practical sample frame and then describe why their sample frame is as good as can feasibly be done.
3. While on-line surveys are appealing because they represent modern technology and are easy, quick and less-costly than mail surveys, their use should be thought through carefully. That "... mail and phone surveys are used more often in survey research, despite the advantages of on-line methods...", [Wade and Parent, 2001], suggests that this guideline is being implemented. Tan and Teo [2000] provide a detailed comparison of mail surveys and Internet surveys which may provide a basis for such careful consideration.
4. In Internet surveys that are not directed at individuals, such as those that are advertised through newsgroups, it is impossible to determine coverage error or nonresponse error. Therefore, such surveys, and all others in which such determinations are not possible, should be avoided unless this method conforms to the special objectives of a study.
5. When nonresponse analyses are performed, at the very least, the researcher should describe the method used and the variables that were analyzed. Preferably, analysis and data would also be provided.
6. Instrument design should be given great attention. The willingness of individuals, particularly managers and professionals, to respond to surveys is an asset that the field should not squander. Surveys that are not carefully designed deplete that asset whereas those that are carefully designed potentially enhance it.

7. Duplicate mail survey instruments should be sent to all sampled individuals who do not respond in a reasonable time. Follow-up reminders (e.g. postcards) should also be considered. Care should be taken not to make such follow-ups when an individual is identified as an "active" non-respondent (e.g. when an instrument is returned that does not entail complete responses or when a no-response policy is articulated).
8. The promise, and delivery, of an "executive summary" of the results is a positive response motivator if the instrument is well-founded. Care must be taken to provide such a summary in a timely fashion, as real-world managers are not used to, and do not understand, the somewhat-more-leisurely pace of academic research as compared with their own work pace. Often a "preliminary summary" that is distributed prior to the completion of data analysis is viewed positively, since practitioners may be more interested in descriptive data than they are in the results of hypothesis testing. In doctoral dissertation research, a summary should be prepared and distributed before, not after, completion of the thesis document.
9. Securing prior management support, or prior acquiescence of potential respondents, or sponsorship of a professional or trade association, provides a convenient and useful point of entry in most organizations. Any or all of these approaches should be considered prior to conducting a survey. In any case, cold-call mailings should be avoided.
10. The goal in the development of every IS survey instrument should be not only to motivate responses but, for example, to create a level of interest so that the respondent requests more information or asks to use the instrument for other purposes. This procedure not only benefit the researcher but also increases the field's asset base.

Acceptance and implementation of these recommendations involves both costs and benefits. More time would be required to do studies; but better, more credible results would be achieved. Fewer survey papers might be published, but the average quality of published papers would be better. Overall, the stature and reputation of IS as a field of research and inquiry might be more positively viewed by other academics and by practitioners.

Editor's Note: This article was received on July 7, 2005 and was published on December 16, 2005. It was with the authors for one revision.

REFERENCES

Ajzen, I (1991) "The Theory of Planned Behavior", *Organizational Behavior and Human Decision Processes*, (50)2, pp. 179-211.

American Association for Public Opinion Research (2000) *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys*, Ann Arbor, MI: AAPOR.

American Association for Public Opinion Research (2004) *Guidelines and Tools for Documenting Survey Results*, Ann Arbor, MI: AAPOR.

Armstrong, C., and V. Sambamurthy (1999) Information Technology Assimilation in Firms: The Influence of Senior Leadership and IT Infrastructures, *Information Systems Research*, (10)4, pp. 304–32.

Bechky, B. (2003) "Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor", *Organization Science* (14)3, pp. 312-331.

Calder, B., L. Phillips, and A. Tybout (1981) "Designing Research for Application", *Journal of Consumer Research*, (8)2, pp. 197-207.

Campbell, D. T. and J. C. Stanley (1966) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Chang, J.C., and W.R. King (2005) "Measuring the Performance of the Information Systems: A Functional Scorecard", *Journal of Management Information System*, (22)1, pp. 85-115.

Chin, W., B. Marcolin, and P. Newsted (2003) "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study", *Information Systems Research*, (14)2, pp. 189-218.

Cook, T.D., and D.T. Campbell (1979) *Quasi-experimentation: Design & Analysis for Field Settings*. Chicago: Rand McNally.

Davis, F. D., R. P. Bagozzi, and P. R. Warshaw (1989) "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models", *Management Science*, (35)8, pp. 982-1003.

Dillman, D. A. (2000) *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley and Sons.

Dooley, L.M. and J.R. Lindner (2003) "The Handling of Nonresponse Error", *Human Resource Development Quarterly,* (14)1, pp. 99-110.

Earl, M. (2001) "Knowledge Management Strategies", *Journal of Management Information Systems*, (18)1, pp. 215-233.

Ellis, R.A., C.M. Endo, and J.M. Armer (1970) "The Use of Potential Nonrespondents for Studying Nonresponse Bias", *Pacific Sociological Review*, (13)1, pp. 103-109.

Fishbein, M. and I. Ajzen (1975) *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Gallupe, R.B., G. DeSanctis, and G.W. Dickson (1988) "Computer-Based Support for Group Problem-Finding: An Experimental Investigation", *MIS Quarterly*, (12)2, pp. 276-297.

Gannon, M., J. Northern, and S. Carroll (1971) "Characteristics of Nonrespondents Among Workers", *Journal of Applied Psychology*, (55)6, pp. 586-588.

Groves R.M., and L.E. Lyberg (1988) "An Overview of Nonresponse Issues in Telephone Surveys", *Telephone Survey Methodology*. New York,: John Wiley & Sons, pp. 191-211.

King, W.R., P. Marks, and S. McCoy (2002) "The Most Important Issues in Knowledge Management", *Communications of the ACM*, (45)9, pp. 93-97.

Lee, A.S., and R.L. Baskerville (2003) "Generalizing Generalizability in Information Systems Research", *Information Systems Research*, (14)3, pp. 221-243.

Linder, J.R., T.H. Murphy, and G.H. Briers (2001) "Handling Nonresponse in Social Science Research", *Journal of Agricultural Education*, (42)4, pp. 43-53.

Luong, A., and S.G. Rogelberg (1998) "How to Increase Your Survey Response Rate", *The Industrial-Organizational Psychologist*, (36)1, pp. 61-65.

Malhotra, M.K., and V. Grover (1998) "An Assessment of Survey Research in POM: From Constructs to Theory", *Journal of Operations Management*, (16)4, pp. 407-425.

McLean, G.N. (1996) "Refereed Journals: The Cornerstone of a Developing Profession", In E. Holton (Ed.) *Proceedings of the Academy of Human Resource Development Annual Conference*, Minneapolis, pp. 360-364.

Nidumolu, S.R., and M.R. Subramani (2003) "The Matrix of Control: Combining Process and Structure Approaches to Managing Software Development", *Journal of Management Information Systems*, (20)3, pp. 159-197.

Orlikowski, W.J. (2002) "Knowing In Practice: Enacting A Collective Capability In Distributed Organizing", *Organization Science* (13)3, pp. 249-274.

Palmer, J.W. (2002) "Web Site Usability, Design, and Performance Metrics", *Information Systems Research*, (13)2, pp. 151-168.

Palvia, P., E. Mao, V. Midha, P. Pinjani, and A.F. Salam (2004) "Research Methodologies in MIS: An Update", *Communications of the Association for Information Systems*, (14)24, pp. 526-542.

Pinsonneault, A., and K.L. Kraemer (1993) "Survey Research Methodology in Management Information Systems: An Assessment", *Journal of Management Information Systems*, (10)2, pp. 75-105.

Richardson, A.J. (2000) "Behaviors Mechanisms of Non-response in Mailback Travel Surveys", *Paper presented at the Seventy-Ninth Meeting of the Transportation Research Board*, Washington, DC.

Rogelberg, S.G, M.E. Sederburg, S. Aziz, J.M. Conway, C. Spitzmüller, and W.E. Knight (2003) "Profiling Active and Passive Nonrespondents to an Organizational Survey", *Journal of Applied Psychology,* (88)6, pp. 1104-1115.

Rogers, E.M. (1983) *Diffusion of Innovations.* Third Edition, New York: Free Press.

Rosenthal, R., and R. Rosnow (1969) *Artifact in Behavioral Research.* New York: Academic Press, pp. 61-112.

Sethi, V., and W.R. King (1991) "Construct Measurement in Information Systems Research", *Decision Sciences*, (22)3, pp. 455-472.

Smith, T.W. (2002) "Reporting Survey Nonresponse in Academic Journals", *International Journal of Public Opinion Research*, (14)4, pp. 469-474.

Sosdian, C.P., and L.M. Sharp (1980) "Nonresponse in Mail Surveys: Access Failure or Respondent Resistance", *Public Opinion Quarterly*, (44)3, pp. 396-402.

Susarla, A., A. Barua, and A.B. Whinston (2003) "Understanding The Service Component Of Application Service Provision: An Empirical Analysis Of Satisfaction With ASP Services", *MIS Quarterly*, (27)1, pp. 91-124.

Swanson, E.B. (1994) "Information Systems Innovation Among Organizations", *Management Science*, (40)9, pp. 1069-1092.

Tan, M., and T.S.H. Teo (2000) "Factors Influencing the Adoption of Internet Banking", *Journal of the Association for Information Systems*, (1)5, pp. 1-42.

Thong, J.Y. (1999) "An Integrated Model of Information Systems Adoption in Small Businesses", *Journal of Management Information Systems*, (15)4, pp. 187-214.

Vessey, I., V. Ramesh, and R.L. Glass (2002) "Research in Information Systems: An Empirical Study of Diversity in the Discipline and Its Journals", *Journal of Management Information Systems*, (19)2, pp. 273-285.

Wade, M.R., and M. Parent (2001) "Relationships Between Job Skills and Performance: A Study of Webmasters", *Journal of Management Information Systems*, (18)3, pp. 71-96.

Watson, R.T., G. DeSanctis, and M. Scott (1988) "Using a GDSS to Facilitate Group Consensus: Some Intended and Unintended Consequences", *MIS Quarterly*, (12)3, pp. 462-479.

Yu, J., and H. Cooper (1983) "A Qualitative Review of Research Design Effects on Response Rates to Questionnaires", *Journal of Marketing Research*, (20)1, pp. 36-44.

Zigurs, I., M.S. Poole, and G.L. DeSanctis (1988) "A Study of Influence in Computer-Mediated Group Decision Making", *MIS Quarterly*, (12)4, pp. 625-644.

ABOUT THE AUTHORS

William R. King holds the title University Professor in the Katz Graduate School of Business at the University of Pittsburgh. He published more than 300 papers and 15 books in Information Systems, Management Science, and Strategic Planning. He served as Founding President of the

Association for Information Systems, President of TIMS (now INFORMS) and Editor-in-Chief of the MIS Quarterly.

Jun He

Jun He is a PhD Student in MIS at the Katz Graduate School of Business at the University of Pittsburgh. He is the coauthor of CAIS Volume 16, Article 32 on Meta Analysis. He presented a number of papers at meetings of the Association for Computing Machinery (ACM) and the Americas' Conference on Information Systems (AMCIS) and participated in the Doctoral Consortium at AMCIS.

Communications of the Association for Information Systems

ISSN: 1529-3181