Validating IS Positivist Instrumentation: 1997-2001

Marie-Claude Boudreau*

Thilini Ariyachandra*

David Gefen**

Detmar W. Straub***

*University of Georgia
MIS Department, Terry College of Business
Athens, GA 30602

**Drexel University
Management Department, Lebow College of Business and Administration
Philadelphia, PA 19104

***Georgia State University
Department of CIS
Robinson College of Business
Atlanta, GA 30302

WORKING PAPER VERSION

Published as: Boudreau, M.-C., Ariyachandra, T., Gefen, D., and Straub, D. "Validating IS Positivist Instrumentation: 1997-2001," in: *The Handbook of Information Systems Research,* M.E. Whitman and A.B. Woszczynski (eds.), Idea Group Publishing, Hershey, PA USA, 2004, pp. 15-26.

Validating IS Positivist Instrumentation: 1997-2001

Abstract

Research quality depends on appropriate statistical validity, a major aspect of which is appropriate instrumentation. This survey examines the quality of IS instrumentation in five leading journals and conclude that much has been done to improve IS research quality through better instrument validation since the last two benchmark studies. Additional recommendations dealing with the current weak spots are discussed.

Keywords: IS research methods; measurement; psychometrics; validation; reliability; content validity; construct validity; manipulation validity; quantitative research; positivist research, guidelines.

INTRODUCTION

The issue of rigor has been the subject of much discussion among information systems (IS) scholars. It has been argued that IS lacks the distinctiveness and rigor usually associated with scientific disciplines and remains institutionally weak (Avgerou, 2000). A specific area where rigor should be improved is the extent of instrument validation, as demonstrated in the work conducted by Straub and his colleagues (e.g., Boudreau *et al.*, 2001; Straub, 1989). In Straub's (1989) article, it was reported that 19% of the articles in three IS journals over a three-year period had utilized either a pretest or a pilot test; that 17% had reported reliability of their scales; and that 14% had validated their constructs, and that only 4% had assessed content validity. These disappointing findings compelled Straub to issue a call for rigorous instrument validation, which was reassessed about a decade later. Boudreau *et al.*'s (2001) study, which expanded the number of sampled articles through the inclusion of additional journals, determined that "some real progress has been made in validating IS research." Indeed, their study showed that 47% of the sampled articles used a pretest or a pilot test; that 63% reported reliability, that 37% validated their constructs; and that 23% assessed content validity. Although such improvements are considerable, Boudreau and her colleagues believed that these percentages were insufficient, and that "[…] the field still has ground to make up to reach more comfortable levels of validation."

Enhancing instrument validation, we argue, will improve the overall process of conducting quantitative research, because it is an elementary building block of statistical validity without which the results of any research are questionable (Cook and Campbell, 1979). In that spirit, this study seeks to provide an up-to-date assessment of the extent to which instrument validation is done rigorously. Our goal is to verify if IS researchers, considering their most recent publications, better validate their research instruments than they did before. We believe that, in the past two years, many researchers and journal editors have responded to the challenge of rigor in instrument validation as they now better understand its importance. This up-to-date assessment considers the same five journals as in Boudreau *et al.*'s (2001) work, but extends the period of coverage to include five full years (i.e., from January 1997 to December 2001). The findings resulting from this inquiry lead us to suggest new guidelines (or reiterate old ones) to be considered by IS researchers, reviewers, and journal editors. The basic premise of this evaluation is that the quality of research design directly manifests itself in the importance of the research findings, and that without good design the conclusions may be unwarranted.

METHOD

The method used in the current research replicates the one used in Boudreau *et al.* (2001). Accordingly, articles were sampled from the same five journals: *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Management Science*, and *Information & Management*. Although the original journal selection was mainly based on Nord and Nord's (1995) study, it is consistent with more recent rankings (i.e., Mylonopoulos and Theoharakis, 2001; Whitman *et al.*, 1999), which consider these five publishing outlets as being important ones within the field of MIS. Given this set of journals — most of which indubitably represent the top tier journals publishing our most important work — the findings of this study will, if anything, *overestimate* the extent to which the field is pursuing appropriate validation efforts. This is appropriate in that we need to see what the best case scenario might be in order to determine whether the quest for validation needs to be reinvigorated at all tiers of journals, or not. If the study raises issues for the top tier journals, then there will surely be validation problems in lower tier journals.

Sampling and Coding Procedures

Articles from these five journals were reviewed, read, and coded, for a period of inquiry starting in January 1997 and ending in December 2001. As in Straub (1989) and in Boudreau *et al.* (2001), the qualifying criteria for the sample was that the article employed either: (a) correlational or statistical manipulation of variables or (b) some form of quantitative data analysis, even if the data analysis was simply descriptive statistics. Studies utilizing archival data (e.g., citation analysis) or unobtrusive measures (e.g., computer system accounting measures) were omitted from the sample unless it was clear from the methodological description that key variable relationships being studied could have been submitted to validation procedures.

Eleven attributes were coded for each surveyed article. First, the *type of research* was assessed, that is, whether the article was confirmatory or exploratory research. As specified by Hair, *et al.* (1995), confirmatory studies are those seeking to test (confirm) a prespecified relationship. Exploratory studies are those that define possible relationships in only the most general form and then allow multivariate techniques to estimate a relationship(s). In the latter case, the researcher is not looking to "confirm" any relationships specified prior to the analysis, but instead allows the method and the data to define the nature of the relationships.

Another coded attribute was the *research method*. Consistent with Stone (1979), Alavi and Carlson (1992), and Alavi, Carlson and Brooke (1989), articles were classified into one of four research methods: (1) laboratory experiments, (2) field experiments, (3) field studies, and

(4) case studies. We relied on Stone (1979) and Stone (1978) for specific definitions and applications of each of these.

Laboratory experiments take place in a setting especially created by the researcher for the investigation of the phenomenon. With this research method, the researcher has control over the independent variable(s) and the random assignment of research participants to various treatment and non-treatment conditions. Field experiments involve the experimental manipulation of one or more variables within a naturally occurring system and subsequent measurement of the impact of the manipulation on one or more dependent variables.

With respect to field studies, they are non-experimental inquiries occurring in natural systems. Researchers using field studies cannot manipulate independent variables or control the influence of confounding variables. For data-gathering techniques, field studies can employ either questionnaires, administered in person, by mail or email, or over the Web, or they can use interview transcripts, coded for quantitative analysis, or they can use a variety of other techniques. Sometimes researchers will refer to “multiple” case studies, which, when they exceed a dozen or more sites, are more than likely classifiable as field studies.

Finally, case studies involve the intense examination of a small number of entities by the researcher, where no independent variables are manipulated nor confounding variables controlled. Like field studies, case studies typically utilize questionnaires, coded interviews, or systematic observation as their preferred techniques for gathering data. Unlike field studies, the foremost concern in case studies is to generate knowledge of the particular (Stake, 1995), from which analytic generalization is possible, rather than statistical generalization (Stake, 1995; Yin, 1994). By intensively studying a small number of entities, a case researcher is likely to develop deep insights of a phenomenon, from which hypotheses may be generated (Yin, 1994).

Benbasat, Goldstein, and Mead (1987) emphasized that a fundamental difference between case studies and the three preceding research methods is that case study researchers generally have less *a priori* knowledge of what the variables of interest will be and how they will be measured.

The extent to which the instrument was refined by a *pretest* and/or a *pilot* study was also coded. As defined by Alreck and Settle (1995), a pretest is a preliminary trial of some or all aspects of the instrument to ensure that there are no unanticipated difficulties. In agreement with Fowler (1984), we contend that every instrument should be pretested, no matter how skilled the researcher, and, therefore, it was critical to capture this attribute.[1] This is necessary even when existing verified scales are applied because different models and settings require revalidation.

As for pilot studies, Alreck and Settle (1995) define them as brief preliminary surveys, often using a small, convenience sample. Pilots test the instrumentation before the project details are finalized and the larger, final survey administered. Another way to distinguish between pilot and pretest is offered by Moser (1958): “the pilot survey is the dress rehearsal, and like a theatrical dress rehearsal, it will be been preceded by a series of preliminary tests and trials [i.e., the pretests]." By gathering information about pilot and pretest, we could assess whether their use has been taken more seriously by the IS research community in the recent years.

Content validity was another attribute collected and coded. Content validity is the degree to which items in an instrument reflect the content universe to which the instrument will be generalized (Cronbach, 1971; Rogers, 1995). This validity is generally established through literature reviews and expert judges or panels. Empirical assessment of content validity is infrequently performed, although Lawshe (1975) provides a procedure and statistic for testing

---

[1] Note that each instrument was treated as a whole. Therefore, in the few cases where a pretest was done on at least one construct within an instrument, then the instrument, in its entirety, was considered as having been pre-tested. The same was true for all criteria that could have been applied differently among constructs. Given this procedure,

this validity. Consistent with Straub (1989), we felt that the sampled domain or content of instruments was an important, but relatively unaddressed area of validation. Measuring the extent to which IS researchers are aware of and implementing content validation was useful information, therefore.

We also tallied articles assessing *construct validity*, that is, the extent to which an operationalization measures the concepts that it purports to measure (Zaltman, Duncan, and Holbek, 1973; Straub, 1989). The focus in construct validity is on whether the selected items "move" together in such a way that they can be considered as an intellectual whole. In establishing construct validity, the researcher is trying to rule out the possibility that constructs, which are artificial, intellectual constructions not directly observable in nature (i.e., “latent”), are being captured by the choices in the measurement instrumentation. Convergent, discriminant, and nomological validation are all considered to be components of construct validity (Bagozzi, 1980). Moreover, criterion-related validity and its sub-types, predictive and concurrent validity (Cronbach, 1990; Rogers, 1995) are also considered to be constituents of construct validity[2].

Establishment of the *reliability* of an instrument, or the absence of the same, was also tallied. As pointed out in Rogers (1995), reliability is a statement about measurement accuracy, i.e., "the extent to which an instrument produces consistent or error-free results.” There are five generally recognized techniques used to assess reliability: (1) internal consistency, (2) split halves, (3) test-retest, (4) alternative or equivalent forms, and (5) inter-rater reliability. In addition, Structured Equation Modeling (SEM) techniques also allow the assessment another special type of reliability that could not be assessed by pre-SEM methods, namely

---

the findings of this study will tend to *overestimate* the extent to which the field is pursuing appropriate validation efforts.

unidimensionality, which assesses whether there is only one underlying construct within each latent construct (Segars, 1997; Gerbing and Anderson, 1988). All of those techniques contribute to establishing the reliability of an instrument. Collecting information about reliability allowed us to determine whether use of the technique is more (or less) frequent than before.

Because *manipulation checks* are critical tests of instrumentation that should always be included when talking about validation principles, this attribute was also coded. Manipulation checks are a variation on tests of construct validity (Perdue and Summers, 1986), and, as such, there can be no assurance of the internal validity of the experiment (Perdue and Summers, 1986) without a verification that the manipulations have "taken." Manipulation checks measure the extent to which treatments have been perceived by the subjects (Bagozzi, 1977). Traditionally, they are incorporated into laboratory and field experiments. It needs to be understood that subjects must be aware of certain aspects of their manipulation, but not others. Manipulation checks are designed to ensure that subjects have, indeed, been manipulated as intended, a validity that can be empirically determined.

For each sampled article, the *nature of the instrument* was coded. An instrument was classified as inaugural, i.e., having been "developed from scratch," or "not based on a previous instrument." Alternatively, an instrument could be based on a previous instrument, in its original version or, more commonly, an adapted version. If no information was provided as to the source of the indicators on the instrument, it was assumed that the research was not based on a preexisting instrument. This attribute was collected to verify if researchers using a preexisting

---

[2] It should be noted that some conceptualize predictive validity as separate and distinct from construct validity (e.g., Campbell, 1960; Cronbach, 1990; Bagozzi, 1980). Other methodologists, however, believe that it may be an aspect of construct validity in that successful predictions of links of constructs to variables outside a theoretical domain also validates, in a sense, the robustness of the constructs (Mumford and Stokes, 1992).

instrument were less inclined to validate all aspects of their instruments than researchers developing their own instrument.

Whether an article incorporated an *instrument validation section*, that is, a specific section concerned with the validation of the instrument, was also evaluated and coded. The addition of an instrument validation section to articles was recommended by Straub (1989) and we felt that it was important to see if IS researchers and journal editors and reviewers were valuing instrument validation enough to include it as a section in its own right.

Finally, we also captured the extent to which studies were making use of second generation statistical techniques, that is SEM tools such as LISREL, PLS, EQS, and AMOS. These techniques offer advantages through the analysis of interrelated research questions by modeling the relationships among multiple independent and dependent constructs simultaneously, in a single, systematic, and comprehensive analysis. Not surprisingly, there has been a growing interest of these techniques among IS researchers, among other reasons because instrument validation is more accurate and comprehensive when done with SEM, and, therefore, more prevalent in IS research that adopts second-generation tools.

STUDY RESULTS

Sample

A total of 293 articles were used in the analysis. Among the articles reviewed, 41 originated in *MIS Quarterly*, 38 in *Information Systems Research*, 119 in *Information & Management*, 78 in *Journal of Management Information Systems* and 17 in *Management Science*[3]. Whereas most were field studies (68%), coded works also included laboratory

[3] The overwhelming number of selected articles from *Information & Management* can be explained by its publication frequency, which is monthly rather than quarterly.

experiments (22%), case studies (5%), and field experiments (5%). In only 23% of the sampled articles, students (undergraduate or graduate) filled out the instrument. In the remainder of the articles, workers (sometimes in conjunction with students) were the ones to whom the instrument was administered. As far as the different techniques for data collection, it is worth mentioning that the majority of our sampled studies have collected data through the use of surveys (questionnaires) (85%). The conduct of an interview was the second technique the most used, with 17% of the sampled articles using it. The use of more than one technique to capture data occurred in 31% of the studies.

Validation of Coding

Our instrument was validated to ensure that the coding was reliable. Since the coding of these articles required an evaluation of textual material, the appropriate test is inter-rater reliability (Miles and Huberman, 1994). To generate this validation test statistic, a second, independent coder was used. For the eleven coded attributes, the following percentages of agreement were obtained: *type of research* – 77%; *research method* – 85%; *pretest* – 82%; *pilot test* – 95%; *content validity* – 90%; *construct validity* – 85%; *reliability* – 85%; *manipulation check* – 95%; *nature of the instrument* – 82%; *instrument validation section* – 92%; and *use of second generation statistical technique* – 95%.

A more stringent coefficient of agreement between judges, Cohen's (1960) kappa coefficient, was also calculated. Specifically, "maximum values" of kappa, as a function of observed agreement levels between judges rather than marginal distributions, were determined, as recommended by Umesh, Peterson and Sauber (1989). For all criteria, the average kappa was 0.76, which is above the 0.70 inter-rater minimum reliability recommended by Miles and

Huberman (1994), Landis and Koch (1977), and Bowers and Courtright (1984). Before further analysis was performed on the collected data, disagreements between the coders were reconciled by one of the authors.

Overview of Findings

Table1 clearly shows that, over the past thirteen years, instrument validation has improved in *all* the categories we have assessed. In addition, in two specific categories (pretest/pilot and reliability), the proportion of published studies validating their instruments is now *greater* than the proportion of published studies not validating their instruments. When comparing with Boudreau *et al.* (2001), the most important improvement is for construct validity, which was then assessed in 37% of the studies compare to 45% of the studies today. Improvement up to an additional 5% is noticeable in all other categories. Overall, although improvement in instrument validation is modest when comparing Boudreau *et al.*'s (2001) results to the current study's results, it is still comforting to observe *consistent* increase in the use of all validation techniques.

Reliability is the validation criterion that was the most frequently assessed, when compared to all other validation criteria taken singly, in both previous studies as well as this one. As it was the case in Boudreau *et al.* (2001), a majority of studies assessing reliability of their instruments have done so through the standard coefficient of internal consistency, i.e., Cronbach's α (84%). The second most popular technique to assess reliability was inter-coder tests, which was reported by 15% of the studies that appraised the reliability of their instrument. Moreover, the use of more than one reliability method is still rare, as it was done by only 10% of the studies assessing reliability.

A closer look at the studies that assessed construct validity reveals that diverse approaches were used for that purpose. More specifically, convergent, discriminant, and nomological validity were determined, respectively, in 50%, 58%, and 6% of these studies. As to predictive and concurrent validity, they were reported in 7% and 1.5% of these studies. Construct validity, in itself (and not in one of its five components), was recounted in 80% of the studies that assessed this kind of validity.

Inst. Categories / Year	**Straub (1989)**	**Boudreau *et al.* (2001)**	**Current Study**
Pretest	13%	26%	31%
Pilot	6%	31%	31%
Pretest or Pilot [4]	19%	47%	51%
Previous Instr. Utilized	17%	42%	43%
Content Validity	4%	23%	26%
Construct Validity	14%	37%	45%
Reliability	17%	63%	68%

Table 1. Survey of Instrument Validation Use in MIS Literature

Table 1 also shows that the utilization of previously existing instruments has more than doubled over the last thirteen years. Also, as detailed in Table 2, it appears that studies using existing instruments were sometimes more inclined to validate their instrument than studies developing their own instrument from scratch. Indeed, construct validity and reliability were more frequently assessed in studies using a previously utilized instrument than those that did not (50% vs. 42%; 74% vs. 63%). However, with regard to the use of pretest or pilot studies and content validity, these validities were assessed more often within studies creating a new instrument than within studies using an existing instrument (55% vs. 46%; 28% vs. 24%). This table reveals another interesting fact: over the past two years, research articles that created their own instrument improved their validation practices to a greater extent than research articles that used a previously utilized instrument.

	Previous Instrument (n=127)		New Instrument (n=166)	
Inst. Categories	**Boudreau *et al.* (2001)**	**Current Study**	**Boudreau *et al.* (2001)**	**Current Study**
Pretest or Pilot	43%	46%	50%	55%
Content Validity	20%	24%	25%	28%
Construct Validity	44%	50%	32%	42%
Reliability	74%	74%	54%	63%

Table 2. Studies with Previously Utilized Instrument versus Those with New Instruments*

Another telling point is how confirmatory studies (133 articles or 45% of total) compare to exploratory studies (160 articles or 55% of total). The present survey indicates that, for all criteria except for the use of pretest or pilot studies, exploratory studies showed less interest in validating their instruments than confirmatory studies (ref. Table 3). Indeed, the extent to which content validity, construct validity and reliability were assessed was more frequent among confirmatory studies than among exploratory studies. This represents the same trend that was observed in Boudreau *et al.* (2001).

	Confirmatory Studies (45%)		Exploratory Studies (55%)	
Inst. Categories	**Boudreau *et al.* (2001)**	**Current Study**	**Boudreau *et al.* (2001)**	**Current Study**
Pretest or Pilot	47%	49%	47%	53%
Content Validity	35%	35%	17%	19%
Construct Validity	53%	61%	29%	33%
Reliability	69%	75%	60%	62%

Table 3. Type of Research (Confirmatory vs. Exploratory Studies)

[4] Note that "Pretest or Pilot" does not add up to "Pretest" plus "Pilot" because some articles used both a pretest and a pilot.

The extent to which a research method has bearing on instrument validation constitutes an interesting observation. In Straub's (1989) original study, it was argued that experimental and case researchers were less prone to validate their instruments than field study researchers. Boudreau *et al.*'s (2001) study showed a similar trend when comparing field studies to experimental studies, but not to case studies. The additional data used in the present study demonstrates that Straub's initial inference still holds true today, on *all* of the previously introduced validity criteria (ref. Table 4). Indeed, field study researchers from our sample were more inclined to validate their instrument than experimental and case researchers. The most notable difference was for the use of construct validity, where a gap of 31% existed between experimental and field study research.

Inst. Categories	**Field Studies (n=200)**	**Lab / Fields Experiments (n=80)**	**Case Studies (n=13)**
Pretest or Pilot	59%	36%	31%
Previous Inst. Utilized	47%	38%	23%
Content Validity	32%	15%	15%
Construct Validity	55%	24%	38%
Reliability	69%	65%	62%

Table 4. Field Studies vs. Lab/Field Experiments vs. Case Studies

The inclusion of an "Instrument Validation" section, as originally suggested in Straub (1989), was tallied as frequently in the current study as it was in Boudreau *et al.*'s (2001) study. Indeed, only 24% of the surveyed articles included such a section. For this minority of articles, there was a greater extent of reporting a pilot or pretest study (80% versus 42%), content validity (52% versus 18%), construct validity (82% versus 34%), and reliability (88% versus 61%). These percentages are hardly surprising as if one feels compelled to include a specific section on instrument validation, it is because efforts have been done in this area. However, it is disappointing not to observe an increase in the percentage of studies that included a special section reporting their endeavor in instrument validation.

Noticeable improvement has occurred in the use of manipulation check in the past few years. As indicated in Table 5, among the field and laboratory experiments in our sample, 30% performed one or several manipulation checks of the treatments, compare to 22% in Boudreau's *et al.*'s (2001) study. Moreover, percentages have particularly increased in two journals, that is, *MIS Quarterly* (increased of 21%) and *Information Systems Research* (increased of 12%). The absence of manipulation checks in the experimental studies of *Management Science* may be due to the tendency, for articles of this journal, to use directly observable measurements, such as time, rather then latent constructs.

Journal	**Boudreau *et al.* (2001)**	**Current study**
Information & Management	24%	25%
Information Systems Research	38%	50%
MIS Quarterly	29%	50%
Journal of Management Information Systems	17%	19%
Management Science	0%	0%
All Five Journals	22%	30%

Table 5. Use of Manipulation Validity

A greater percentage of studies from our sample used second generation statistical techniques (e.g., structural equation modeling) rather than first generation statistical techniques (regression, ANOVA, LOGIT, etc.). From 15% in Boudreau *et al.* (2001), this percentage increased to 19% in the present study (ref. Table 6). However, the extent of instrument validation did not change much when comparing first to second generation techniques in the two studies. As it was the case in Boudreau *et al.* (2001), studies making use of SEM techniques scored higher in all categories, particularly for construct validity and reliability. Among the studies using second-generation statistical techniques, the most commonly used tools were PLS (42%), LISREL (21%), and EQS (18%).

A possible reason for this difference is that SEM analyzes both the *structural model* (the assumed causation) and the *measurement model* (the loadings of observed items). As a result, validity assessment is an integral part of SEM. The validity statistics appear explicitly in the output, and the degree of statistical validity directly affects the overall model fit indexes. In first generation statistical techniques, on the other hand, validity and reliability are performed in separate analyses that are not related to the actual hypothesis testing and, thus, do not determine the overall fit indexes.

	Boudreau *et al.* (2001)		**Current study**	
Inst. Categories	**First Generation (85%)**	**Second Generation (15%)**	**First Generation (81%)**	**Second Generation (19%)**
Pretest or Pilot	44%	64%	48%	63%
Previous Inst. Utilized	42%	46%	43%	46%
Content Validity	19%	43%	23%	39%
Construct Validity	29%	82%	36%	86%
Reliability	57%	96%	61%	93%

Table 6. First Generation vs. Second Generation Statistical Techniques

Summary of Key Points

It should be considered good news that, in the short period of two years since the last study assessing instrument validation practices, IS researchers have improved the validation of their instrument. Granted, such an improvement is certainly not as significant as what had been observed when using Straub's (1989) study as the baseline, but this is understandable given that the time period was then much longer. Although better, the current validation practices are far from perfect, and it is still necessary to state that IS researchers need to achieve greater rigor in the validation of their instruments and their research. More particularly, the following nine key findings should engage further reflection and action:

1. Over the past two years, instrument validation practices have steadily improved.
2. In two specific categories (pretest/pilot and reliability), the proportion of published studies validating their instruments is now *greater* than the proportion of published studies not validating their instruments.
3. The assessment of construct validity has improved the most over the past two years.
4. Published studies are increasingly using pre-existing instruments; while doing so, reliability and construct validity are being more frequently assessed.
5. Confirmatory studies are more likely to assess reliability, content validity and construct validity than exploratory studies.
6. Laboratory experiments, field experiments, and case studies, lag behind field studies with respect to all validation criteria.
7. Although the inclusion of an "Instrument Validation" sub-section warrants greater reporting of validation practices, it appears infrequently in empirical studies.
8. There has been a noticeable improvement in the use of manipulation check in the past few years; however, in some publications outlets, manipulation check in only done by a minority of IS experimenters.
9. Published studies making use of second generation statistical techniques (SEM) are much more likely to validate their instruments than published studies making use of first generation statistical techniques.

DISCUSSION

Research quality is a never-ending process of striving for improvement and of a process of reexamination of previous methods. It is a crucial process, for it is through this constant cycle of self-assessment and consequential improvement that scientific methods are improved and that the results of science become more reliable, valid, and significant. We are glad to report that this self-improvement process seems to be working in IS research. The study shows there is reason to be optimistic, although much still needs to be done. In all the aspects we examined the results of this study show that there has been remarkable improvement in instrument validation practices since Straub's initial plea in 1989. Furthermore, even in the short time period of two years that differentiates this study from Boudreau *et al.*'s (2001) study, it is clear that validation practices have improved some, at least modestly. This seems to be at least in part because of the growing awareness to the need to ensure quality research in the IS community and the growing popularity of SEM tools (see Gefen *et al.* (2000) for a comparison of these). It appears that researchers, reviewers, and journal editors, have made a conscious effort in the past years to include more rigor in their work. This should, if the trend continues, help reduce the skeptical questions about the quality of the work conducted by IS researchers (e.g., Avgerou, 2000).

Looking at the half empty part of the glass, we could also observe that instrument validation practices are yet to be carried out a majority of the time, except for the assessment of reliability and the use of a pretest/pilot study. In addition, one should realize that these results are overestimates (as explained earlier), and that instrument validation is likely to be done to a much lesser extent in less prestigious IS journals and conferences. In other words, one can surmise that in many cases, research is being conducted (and later published) with little validation of the data collection instruments. What does that say about the quality of our work? We caution the IS research community not to have a false sense of complacency when looking at the results of the current study. Improvement has been noticed, yes, but much more lies ahead and needs to be done.

Perhaps one of the most noticeable changes has been in the increased application of SEM and with it the heightened attention researchers are forced to give to verifying construct validity and reliability. Where in first generation tools researchers could figuratively sweep unwarranted variance under the carpet, with SEM researchers are forced to a much larger degree to face this unwarranted variance with SEM tools and account for it by improving their scales and model. Clearly, statistical analysis when done with an appropriate SEM tool is much better, more reliable, and shows greater construct validity (Hair *et al.*, 1995; Gefen, *et al.*, 2000). It is, in fact, hard to get a good structural model in SEM where the measurement model statistics, namely reliability and construct validity, are not up to the standard. Indeed, the reporting of these two statistics is considerably higher with research based on SEM as its analysis tool.

CONCLUSION

Good research is built first and foremost on good research design and correct data analysis. Science will not progress without statistical validity that is brought about by a meticulous attention to instrument validation through pretesting and pilot testing, to reliability, to content and construct validity, and to the correct application of appropriate tools. With this in mind, the results of this survey are encouraging.

References

Alavi, M., and Carlson, P. "A Review of MIS Research and Disciplinary Development," *Journal of Management Information Systems* (8:4), Spring 1992, pp. 45-62.

Alavi, M., Carlson, P., and Brooke, G. "The Ecology of MIS Research: A Twenty Year Status Review," *Proceedings of the Tenth International Conference on Information Systems,* Boston, MA, 1989, pp. 363-271.

Alreck**,** P.A., and Settle, R.B. *The Survey Research Handbook*, Irwin, Chicago, Second Edition, 1995**.**

Avgerou, C., "Information systems: What sort of science is it?," *Omega*, (28:5), 2000.

Bagozzi, R.P. "Structural Equation Models in Experimental Research," *Journal of Marketing Research* (14), 1977, pp. 209-236.

Bagozzi, R.P. *Causal Methods in Marketing*, John Wiley and Sons, New York, 1980.

Benbasat, I., Goldstein, D.K., and Mead, M. "The Case Research Strategy in Studies of Information Systems," *MIS Quarterly* (11:3), 1987, pp. 369-388.

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

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

Campbell, D.T. "Recommendations for APA Test Standards Regarding Construct, Trait, Discriminant Validity," *American Psychologist* (15), August 1960, pp. 546-553.

Cohen, J. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20), 1960, pp. 37-46.

Cook, T.D., and Campbell, D.T. *Quasi Experimentation: Design and Analytical Issues for Field Settings*, Rand McNally, Chicago, 1979.

Cronbach, L.J. "Test Validation," In *Educational Measurement*, R. L. Thorndike (Ed.), American Council on Education, Washington, D.C., 1971, pp. 443-507.

Cronbach, L.J. *Essentials of Psychological Testing*, Harper-Row, New York, 1990.

Fowler, F.J., Jr, *Survey Research Methods*, Sage Publications, Beverly Hills, CA, 1984.

Gefen, D., Straub, D., and Boudreau, M.-C. (2000). "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of AIS* (4:7).

Gerbing, D.W. and Anderson, J.C. (1988). "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, Vol. 25, No. May, pp. 186-192.

Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., and Black, W.C. *Multivariate Data Analysis with Readings*, Prentice Hall, Englewood Cliffs, NJ, 1995.

Landis, J.R., and Koch, G.G. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (22), 1977, pp. 79-94.

Lawshe, C.H. "A Quantitative Approach to Content Validity," *Personnel Psychology* (28), 1975, pp. 563-575.

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

Moser, C.A. *Survey Methods in Social Investigation*, Heinemann, London, 1958.

Mumford, M.D., and Stokes, G.S. "Developmental Determinants of Individual Action: Theory and Practice in the Application of Background Data Measures," In *Handbook of Industrial and Organizational Psychology*, M. D. Dunnette and L. M. Hough (Ed.), 2, Consulting Psychologists Press, Palo Alto, CA USA, 1992, pp. 61-138.

Mylonopoulos, N. and V. Theoharakis, "On-Site: Global Perceptions of IS Journals," *Communications of the ACM*, Sept. 2001, vol. 44, no. 9, pp. 29-33.

Nord, J. H., and Nord, G. D. "MIS Research: Journal Status Assessment and Analysis," *Information & Management* (29:1), 1995, pp. 29-42.

Perdue, B.C., and Summers, J.O. "Checking the Success of Manipulations in Marketing Experiments," *Journal of Marketing Research* (23:4), November 1986, pp. 317-326.

Rogers, T.B. *The Psychological Testing Enterprise*, Brooks/Cole Publishing Company, Pacific Grove, CA, 1995.

Segars, A.H. "Assessing the Unidimensionality of Measurement: A Paradigm and Illustration within the Context of Information Systems Research," *Omega* (25:1), February 1997, pp. 107-121.

Stake, R. E., *The Art of Case Study*. Thousand Oaks, CA: Sage, 1995.

Stone, E.F. "Research Methods and Philosophy of Science," In *Organizational Behavior*, S. Kiev (Ed.), Columbus, Ohio, 1979, pp. 15-40.

Stone, E.F. *Research Methods in Organizational Behavior*, Scott, Foresman & Company, Glenview, IL, 1978.

Straub, D.W. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), 1989, pp. 147-169.

Umesh, U.N., Peterson, R.A., and Sauber, M.H. "Interjudge Agreement and the Maximum Value of Kappa," *Educational and Psychological Measurement* (49), 1989, pp. 835-850.

Whitman, M., Hendrickson, A. and Townsend, A. "Research Commentary. Academic Rewards for Teaching, Research and service: Data and Discourse," *Information Systems Research*, June 1999, vol. 10, no. 2, pp. 99-109.

Yin, R.K. *Case Study Research: Design and Methods*, Sage Publications, Thousand Oaks, CA, 1994.

Zaltman, G., Duncan, R., and Holbek, J. *Innovations and Organizations*, John Wiley, New York, NY, 1973.