

Validation in IS Research: A State-of-the-Art Assessment

Marie-Claude Boudreau*

David Gefen**

Detmar W. Straub***

***University of Georgia**

Management Information System Department
Terry College of Business
University of Georgia
Athens, Georgia 30602

****Drexel University**

Management Department
College of Business and Administration
Drexel University
32nd and Chestnut Streets
Philadelphia, PA 19104

*****Georgia State University**

Department of Computer Information Systems
College of Business Administration
Georgia State University
Atlanta, Georgia 30302-4015

WORKING PAPER VERSION

Published as: Boudreau, M., Gefen, D., and Straub, D. "Validation in IS Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1, March) 2001, pp 1-24.

Validation in IS Research: A State-of-the-Art Assessment

Abstract

Over ten years ago, Straub (1989) raised the issue of whether IS researchers were rigorously validating their quantitative, positivist instruments. In the years that have passed since that time, the profession has undergone many changes. Novel technologies and management trends have come and gone. New professional societies have been formed and grown in prominence and new demands have been placed on the field's research and teaching obligations. But the issue of rigor in IS research has persisted throughout all such changes. Without solid validation of the instruments that are used to gather data upon which findings and interpretations are based, the very scientific basis of positivist, quantitative research is threatened.

As a retrospective on the Straub article, this research seeks to determine if and how the field has advanced in instrument validation. As evidence of the change, we coded positivist, quantitative research articles in five major journals over the last three years for use of validation techniques. Findings suggest that the field has advanced in many areas, but, overall, it appears that a majority of researchers are still *not* sufficiently validating their instruments. Based on these findings, approaches are suggested for reinvigorating the quest for validation in IS research via content/construct validity, reliability, and manipulation validity.

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

“The proof of the pudding is in the eating”
—*Miguel Cervantes (from Don Quixote)*

1. NEED FOR REASSESSMENT OF INSTRUMENT VALIDATION

Information systems (IS) research is a dynamic and ever changing field. Novel technologies and management trends have emerged, evolved, and departed over the years since the discipline first formed in the late 1960's. In recent years, new professional societies have been formed and new demands have been placed on the field's research and teaching obligations. But the problem of rigor in IS research has persisted through all such changes. Within the positivist, quantitative arena of research, the very scientific basis of the profession depends on solid validation of the instruments that are used to gather the data upon which findings and interpretations are based.

In 1989, Straub issued a call for new efforts to validate the positivist, quantitative instruments that IS researchers were using. In a straight-forward coding of the use of various techniques for gathering empirical data, he found that only **17%** of the articles in three widely referenced IS journals over the previous three years had reported reliability of their scales. In addition, only **13%** had validated their constructs, while a scant **19%** had utilized either a pretest or a pilot test. The argument for validation of instruments was based on the prior and primary need to validate instruments before such other crucial validities as internal validity and statistical conclusion validity are considered.

What has changed, if anything, in the intervening years? Have IS researchers conducting positivist, quantitative research responded to the challenge and emphasized instrumentation validity in their work? The question and answer are critical if we are to understand how IS is evolving as a discipline. It is also important as a guide for better practice, both for journal

editors/reviewers and for authors. The current study is a review of the state of the art of validation in IS research, as it pertains to research that is both positivist and quantitative. Our study of research practice over a recent three year period suggests that the field has advanced in a number of essential ways. It also highlights many areas that are lagging and call for greater attention.

2. METHOD

In order to determine whether instruments used by the IS community are more thoroughly validated recently than eleven years ago, a new review and analysis of the literature was conducted. Because we wanted to compare the current situation with what was prevalent at the time of Straub's (1989) research, we maintained certain commonalities with the previous study when it was reasonable to do so. In this spirit, the sampling and coding procedures were similar.

Journal Selection

The three journals sampled in Straub's original piece (*MIS Quarterly*, *Communications of the ACM*, and *Information & Management*) represented journals that were publishing important research in the field in the late 1980's. Because the period of inquiry for the current research covers the three years between 1997 to 1999, journals representative of the research conducted during this more recent era were selected. As a guide in making this selection, we used Nord and Nord's (1995) evaluation of top-ranked journals. Using previous studies as a filter, these authors identified nine journals from the field of Management Information Systems that were considered

to be "top tier."¹ Among these nine journals, Nord and Nord concluded that six published "important" or "very important" research.² A sub-set of four journals among these six were chosen for our sampling: *MIS Quarterly*, *Information & Management*, *Journal of Management Information Systems*, and *Management Science*. To this list, we added a fifth journal, *Information Systems Research*, because it is generally recognized as comparable to *MIS Quarterly* in quality (Hardgrave and Walstrom, 1997; Walstrom et al., 1995). Since it was a rather recent journal (founded in 1991), it is understandable why it was not in Nord and Nord's ranking.

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.³ 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 with 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 1999. As in Straub (1989), 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

¹ In alphabetical order, these journals were: *Communications of the ACM*, *Decision Sciences*, *Information & Management*, *Information Systems Management*, *Journal of Computer Information Systems*, *Journal of Management Information Systems*, *Journal of Systems Management*, *Management Science* and *MIS Quarterly*.

² Nord and Nord (1995) singled out three journals (*Communications of the ACM*, *Information Systems Management* and *Journal of Systems Management*) in which research-based manuscripts were considered less important.

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.

To enrich the analysis, we expanded the attributes collected by Straub (1989). In total, 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 which 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. Type of research, we felt, was important to collect because exploratory studies were expected to be less thorough in their instrument validation than confirmatory studies.

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.

As for 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. Finally, case studies involve the intense examination of one or a few 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 generalization⁴ is possible, rather than statistical generalization (Stake, 1995; Yin, 1994). By intensively studying one or a few 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.⁵ With these four types of research methods, we classified each sampled article.⁶ One reason it was desirable to capture research method type was that, in 1989,

⁴ Analytic generalization, which is also called theoretical elaboration, refers to the study of some phenomenon in a particular set of circumstances to support, contest, refine, or elaborate a theory, model, or concept. (Schwandt, 1997).

⁵ Benbasat, et al. (1987) also emphasized, however, that the lack of *a priori* knowledge is sometimes a matter of degree. An example of an article that fits the case study research method is Webster's (1998) research. In this article, the author investigated the use of videoconferencing in particular organization over a one year period. Although she started with a particular theoretical perspective in mind (i.e., communication media choice), preliminary study results suggested the inclusion of two additional theoretical perspectives (i.e., system analysis and design and privacy). These additional perspectives, in turn, influenced subsequent data collection (through interviews and questionnaires). Because no independent variables were manipulated nor confounding variables controlled, because it had to do with the intense inquiry of a single site, and because not all the appropriate theoretical constructs were decided *a priori* (some emerged *during* the conduct of the study), Webster's study was classified as a case study.

⁶ Stone (1979) also mentions "sample survey" as another type of research method. We prefer to conceptualize sample survey as a *technique* for data collection rather than as a method, that is, a means to systematically collect

experimental research used fewer validated instruments than non-experimental research (Straub, 1989) and we wanted to see if experimentalists had strengthened their methods over the last decade.

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.

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 preceded by a series of preliminary tests and trials [i.e., the pretests]." Pretest and pilot attributes were also collected by Straub in 1989, who recommended a greater use of these. By gathering the same information, we can assess whether use of pre- and pilot tests had been taken more seriously by the IS research community in the intervening 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

data from individuals in a sample. Given this definition, a sample survey (mailed or telephone survey) could be used as data collection technique for either field studies or case studies. Moreover, questionnaires or surveys are also

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 et al., 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 validity⁷.

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: internal consistency, split halves, test-retest, alternative or equivalent forms, and inter-rater reliability. Most recently, these techniques have been supplemented by reliability assessed through Structural Equation Modeling (Segars, 1997), which has been dubbed "unidimensional reliability." All of those techniques contribute to

frequently used in experimentation.

establishing the reliability of an instrument. In that the reliability of an instrument was also collected by Straub in 1989, collecting information about reliability allowed us to determine whether use of the technique is more (or less) frequent today than eleven years ago.

Because *manipulation checks* are critical tests of instrumentation that should always be included when talking about validation principles, this attribute was also coded. Despite the fact that these checks were not covered in Straub (1989), we felt that the IS field needed to begin to gauge the state-of-the-art of IS experimental validation. 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 test our belief that researchers using a

⁷ 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).

preexisting 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 prior to the year 2000 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, Structural Equation Modeling (SEM) tools such as LISREL, PLS, EQS or 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. Instrument validation should be easier to do with SEM, and, therefore, more prevalent in IS research that adopts second generation tools.

3. STUDY RESULTS

Sample

A total of 207 articles were used in the analysis.⁸ Among the articles reviewed, 35 originated in *MIS Quarterly*, 26 in *Information Systems Research*, 84 in *Information & Management*, 50 in *Journal of Management Information Systems* and 12 in *Management*

⁸ Overall, 394 articles were considered for inclusion. More particularly, the distribution of articles was as followed: 143 in *Information & Management*; 65 in *Information Systems Research*; 63 in *MIS Quarterly*; 102 in *Journal of Management Information Systems*; and 21 in *Management Science*. The qualifying criteria, as explained earlier in this paper, excluded 187 articles, overall.

*Science*⁹. Whereas most were field studies (60%), studied works also included field experiments (5%), laboratory experiments (23%) and case studies (12%).

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* – 74%; *research method* – 81%; *pretest* – 81%; *pilot test* – 93%; *content validity* – 89%; *construct validity* – 81%; *reliability* – 85%; *manipulation check* – 96%; *nature of the instrument* – 81%; *instrument validation section* – 89%; and *use of second generation statistical technique* – 100%.

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 0.75, 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

⁹ The overwhelming number of selected articles from *Information & Management* can be explained by its publication frequency, which is monthly rather than quarterly.

As indicated in Table 1, the data clearly shows that researchers are validating their instruments more than they were eleven years ago¹⁰. Nevertheless, it appears that the proportion of researchers validating their instruments is smaller than the proportion of researchers *not* validating their instruments in all categories, except for the use of reliability. Both pretest and pilot studies were more frequently employed, that is, 45% of the articles used one or the other. Such improvement is considerable, given that this figure was only 19% a decade ago. Similarly, content and construct validities, respectively at 4% and 14% in the 1989 study, rose to 26% and 34% in the current survey. With regard to reliability, researchers are utilizing these tests more widely. Whereas 17% of the surveyed studies assessed reliability in 1989, 59% perform these tests now. On this matter, it is worth mentioning that the majority of the studies assessing the reliability of their instrument have done so through the standard coefficient of internal consistency, i.e., Cronbach's α (78%). Only in rare cases have other methods been used to verify reliability of measures. Specifically, 2% used test/retest, 2% used split halves, and 20% used inter-coder tests, such as the validation statistics reported for our study. Moreover, the use of more than one reliability method occurred in 13% of the studies assessing reliability.

It is surprising that the extent of instrument validation is much lower in most categories in the IS articles in *Management Science* than the other journals. While straightforward techniques such as reliability and pretests are extremely low (8% and 8%, respectively), more sophisticated validation like construct validity are likewise comparatively low (25%). Given the preeminence of *Management Science* as one of the most widely respected journals in the business disciplines, this is a startling result. The *Management Science* sample was smaller than

¹⁰ The extent to which validation practices improved over time was also examined. From 1997 to 1999, our sample did *not* reveal a clear tendency toward overall better validation practices. Indeed, the percentage of researchers assessing each of the different validation criteria were approximately the same over the years 1997, 1998 and 1999.

that of the other journals, but sufficient for comparison purposes, so there is no ready explanation for this deficiency.

Inst. Categories	2000						1989
	I&M	ISR	MISQ	JMIS	MS	Total	Total
Pretest	26%	35%	29%	20%	8%	25%	13%
Pilot	31%	19%	29%	34%	17%	29%	6%
Pretest or Pilot ¹¹	45%	46%	43%	50%	25%	45%	19%
Prev. Instr. Utilized	33%	58%	26%	56%	25%	40%	17%
Content Validity	17%	31%	51%	14%	50%	26%	4%
Construct Validity	30%	54%	31%	36%	25%	34%	14%
Reliability	54%	85%	54%	72%	8%	59%	17%

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 decade. Contrary to expectations, studies using existing instruments were sometimes more inclined to validate their instrument than studies developing their own instrument from scratch. Indeed, as detailed in Table 2, construct validity and reliability were more frequently assessed in studies using a previously utilized instrument than those that do not (43% vs. 28%; 73% vs. 50%). 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 (47% vs. 42%; 29% vs. 20%).

¹¹ Note that "Pretest or Pilot" does not add up to "Pretest" plus "Pilot" because some articles used both a pretest and a pilot.

Inst. Categories	Previous Instrument (n=83)		New Instrument (n=124)	
	Tot N	Tot %	Tot N	Tot %
Pretest or Pilot	35	42%	58	47%
Content Validity	17	20%	36	29%
Construct Validity	36	43%	35	28%
Reliability	61	73%	62	50%

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

Another telling point is how confirmatory studies (64 articles or 31% of total) compare to exploratory studies (143 articles or 69% of total). The present survey indicates that, as we suspected, researchers engaged in exploratory studies generally showed less interest in validating their instruments than those engaged in 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. The extent to which a pretest or pilot studies were performed, however, was as frequent within confirmatory studies as it was within exploratory studies.

Inst. Categories	Confirmatory Studies (n=64)		Exploratory Studies (n=143)	
	Tot N	Tot %	Tot N	Tot %
Pretest or Pilot	29	45%	64	45%
Content Validity	24	38%	29	20%
Construct Validity	33	52%	38	27%
Reliability	44	69%	79	55%

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

Based on the results shown in Tables 2 and 3, one might be tempted to conclude that most confirmatory studies from our sample used a previously validated instrument, whereas most exploratory studies created their own instrument. Such a conclusion is accurate, for the most part. On the one hand, the sampled confirmatory studies used an existing instrument almost as

often as they created a new instrument (47% vs. 53%). On the other hand, the sampled exploratory studies had many more researchers developing a new instrument (63%) than using an instrument previously developed (37%). Confirmatory studies, as it turns out, are not only more inclined to validate their instruments, but they also are more likely to reuse an already existing instrument, as shown in Table 4.

Use of an Instrument	Confirmatory Studies (n=64)		Exploratory Studies (n=143)	
	Tot N	Tot %	Tot N	Tot %
Previous Instrument	30	47%	53	37%
New Instrument	34	53%	90	63%

Table 4. Relationship between Type of Research and Use of Existing Instrument

Another relevant factor expected to influence the extent to which an instrument is validated is the research method selected. In his 1989 study, Straub noted that experimental and case researchers were less prone to validate their instruments than field study researchers. The present study shows a similar trend when comparing experimental studies to field studies. Indeed, as Table 5 illustrates, most of the previously introduced validity criteria were proportionally less prominent in laboratory and field experiments than they were in field studies. Only the assessment of reliability is as frequently conducted within experiments as it is within field studies (62%).

However, when comparing case studies to field studies, the story is slightly different. Although field study researchers are more inclined to assess reliability, construct validity, and to use a pretest or pilot test to validate their instrument, they are less likely to assess content validity than are case study researchers.

Inst. Categories	Field Studies (n=124)	Lab/Field Experiments (n=58)	Case Studies (n=25)
Pretest or Pilot	52%	40%	20%
Prev. Inst. Utilized	46%	38%	16%
Content Validity	26%	17%	44%
Construct Validity	44%	21%	20%
Reliability	62%	62%	40%

Table 5. Field Studies vs. Lab/Field Experiments vs. Case Studies

Straub (1989) encouraged researchers to include an "Instrument Validation" subsection when accounting for results. The present research found that only 23% of the surveyed articles did so over the investigated period. Interestingly enough though, this percentage is not uniform across the surveyed journals. Whereas more than 42% of selected articles in *Information Systems Research* had this specific subsection, less than 27% of the selected articles in the other journals had it. More specifically, *MIS Quarterly*, *Management Science*, *Journal of Management Information Systems*, and *Information & Management* had, respectively, 26%, 25%, 20%, and 18% of their sampled articles including this section. It appears that, in this particular way, practices vary considerably across journals.

As Table 6 points out, manipulation validity was infrequently assessed in experimental settings. Among the 58 field and laboratory experiments included in our sample, only 13 (22%) performed one or several manipulation checks of the treatments. Moreover, as was the case for the inclusion of an "Instrument Validation" subsection, percentages vary across the surveyed journals. Indeed, 38% of the experiments in *Information Systems Research* included a manipulation check, whereas the equivalent percentages for *MIS Quarterly* and *Information & Management* were considerably lower, at 29% and 24%, respectively. *Journal of Management Information Systems* and *Management Science* scores were even lower, at 17% and 0%. The

absence of manipulation check in the experimental studies of *Management Science* may be due to the tendency, specifically in the articles of this journal, to use directly observable measurements, such as time, rather than latent constructs. As to the particular means by which manipulation validity was assessed, our sample reveals that more sophisticated statistics (t-test, chi-square, ANOVA) were deployed about twice as often as simpler descriptive statistics (counts, means, percentages).

The relatively low percentage of experimenters that validate the effectiveness of their treatments may be tempered by the fact that in some cases, these manipulation checks could have been performed but not reported.¹² It is also quite possible that reviewers did not require these checks even though the researchers had actually performed them.¹³

Journal	# Lab/Field Experiments	# w/ Manip. Check	Total %
<i>Information & Management</i>	17	4	24%
<i>Information Systems Research</i>	8	3	38%
<i>MIS Quarterly</i>	7	2	29%
<i>Journal of Management Information Systems</i>	24	4	17%
<i>Management Science</i>	2	0	0%
All Five Journals	58	13	22%

Table 6. Use of Manipulation Validity

Do studies using first generation statistical techniques differ from studies using second generation statistical techniques (e.g., Structural Equation Modeling) in terms of their validation practices? As shown in Table 7, this prediction seems to be the case. In all categories investigated, studies making use of SEM techniques scored higher. The extent to which construct validity and reliability are more often assessed within SEM studies is particularly striking.

¹² The article as artifact is imperfect, of course. It may be that many of the observed characteristics in this trend analysis suffer from this same limitation.

¹³ Of course, the same speculation can be made about the other validities, but reviewers are likely more aware of these and, therefore, more likely to ask authors to report these.

Indeed, there is a difference of 55% in the proportions of studies assessing construct validity if we compare SEM to non-SEM studies. Similarly, there is a difference of 42% in the proportions of studies assessing reliability if we compare SEM to non-SEM studies.

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 hypotheses testing and, thus, do not determine the overall fit indexes.

Inst. Categories	First Generation (n=179)	Second Generation (n=28)
Pretest or Pilot	42%	64%
Prev. Inst. Utilized	39%	46%
Content Validity	23%	43%
Construct Validity	27%	82%
Reliability	54%	96%

Table 7. First Generation vs. Second Generation Statistical Techniques

Summary of Key Points

Overall, despite the improvements made over the previous years, much still needs to be done in order for IS researchers, in general, to achieve necessary rigor in the validation of their instruments and their research. More particularly, the following eight key findings should engage further reflection and action:

1. IS researchers validating their instruments have yet to be in the majority in any single category other than in reliability.
2. Reliability is usually limited to Cronbach's α ; alternative methods, or a combination of methods, for validation of mature research streams, are rarely used.

3. IS researchers are increasingly using pre-existing instruments; while doing so, reliability and construct validity are being more frequently assessed.
4. Confirmatory studies are more likely to assess reliability, content validity and construct validity than exploratory studies.
5. Laboratory and field experimentation lag far behind field studies with respect to instrument validation.
6. "Instrument Validation" sub-sections appear infrequently in empirical studies, although this is, to some extent, a function of the publication outlet.
7. A small minority of IS experimenters validate their manipulations.
8. Researchers making use of second generation statistical techniques (SEM) are much more likely to validate their instruments than researchers making use of first generation statistical techniques.

The next section draws upon these key findings to support the point that, instrument validation can, and we will argue, *ought to be given even greater attention* in the immediate future.

4. DISCUSSION

Overall, the data clearly shows that there was an across-the-board improvement in all validation practices. Nevertheless, one may infer equally that, as a field, we are far from laying to rest the issue of instrument validation! The field still has ground to cover and to make more rigorous and credible the instruments used to conduct IS quantitative, positivist research

In 1989, Straub offered the following guidelines for practice:

1. Researchers should pretest and/or pilot test instruments, attempting to assess as many validities as possible in this process.
2. MIS journal editors should encourage or require researchers to prepare an "Instrument Validation" sub-section of the Methodology section; this sub-section can include, at the least, reliability tests and validity tests of the final administered instrument.
3. Researchers should use previously validated instruments wherever possible, being careful not to make significant alterations in the validated instrument without revalidating instrument content, constructs, and reliability.
4. Researchers should undertake formal validation [i.e., multitrait-multimethod analysis, structural equation modeling, and other techniques for thoroughly assessing convergent

and discriminant validity] whenever it is critical to ground central constructs in major MIS research streams.

For the further maturation of the IS field, we have found every reason to believe that these guidelines are as critical in 2000 as they were in 1989. As reported earlier, more IS quantitative, positivist researchers are validating their studies today, but the field has still not reached the point where validation is the rule rather than the exception. In only one category in the present study did the number of researchers engaging in a validation practice exceed those who did not, namely in the practice of reliability. What this slow progress toward rigorously validated instruments suggests is that the guidelines for research practice may need to be strengthened. Exactly how and in what specific ways, i.e., which practices are *mandatory* and which are *recommended*, needs to be dealt with in further conceptualization of methodological practice in IS research.

There is one point in the 1989 guidelines which calls for further clarification, based on practice regarding this specification since the publication of this article. Guideline #3 suggests that previously validated instruments should be used wherever possible, which some researchers may have interpreted to mean that use of previously validated instruments is a *superior* practice to re-validating and/or creating new measures for constructs.

Nothing could be further from the truth.

Use of previously validated instruments is a practice that is efficient, but one that does not lead to better constructs for the field. An important reason for using previously validated instruments is theoretical, that is, it allows researchers to cumulate knowledge and assure comparability between studies. Another reason for choosing to use previously validated instruments is practical. The context of this guideline in Straub (1989) was the following: "In a technology-driven field such as MIS, windows of opportunity for gathering data appear and

disappear so rapidly that researchers often feel they cannot afford the time required to validate their data collection instruments" (p. 161). This situation is still with us today and so we must recognize that the fast pace of technological change creates important challenges for researchers and often prevents them from investing time in developing new instruments.

Nevertheless, some researchers will elect to devote the time and effort to create new scale items rather than use previously validated scales. By so doing, they will, in fact, be exposing the constructs to a more robust test of construct validity (Cook and Campbell, 1979; Sussmann and Robertson, 1986). In other words, if theoretical linkages hold using alternative methods of measuring the constructs, then this would be sufficient evidence for the nomological (construct) validity of both scales. Validation always works in both directions; it is "symmetrical and egalitarian" (Campbell, 1960, p. 548). Creating new instruments from scratch for even well established constructs is not an efficient practice, but it should never be discouraged.

For clarity, therefore, guideline #3 can be reconceptualized as:

3. For the sake of efficiency, quantitative, positivist researchers should use previously validated instruments wherever possible, being careful not to skirt previous validation controversies or to make significant alterations in validated instruments without revalidating instrument content, constructs, and reliability. Researchers who are able to create new instrumentation for established constructs would be testing the robustness of the constructs and theoretical links to method/measurement change (nomological validity). Creating new validated instruments, thus, is highly desirable.

The phrase "being careful not to skirt previous validation controversies" has been added to the guideline for several reasons. While use of previously validated instruments does, in general, save time, there can be drawbacks to its use that reduces this efficiency and requires journal space to explicate. Controversies may be surrounding these prior validations and researchers, editors, and reviewers need to both be aware of these controversies and attempt to deal with them up-front in the article itself.

A good example of this situation is the SERVQUAL instrument, which was introduced into IS research by Pitt, Watson, and Kavan (1995) and was later challenged by Van Dyke, Kappelman, and Prybutok (1997). Whereas Pitt, Watson, and Kavan (1997) responded to this challenge and presented reasons why they feel the instrument is still valid, the IS community would have been well served if the authors had been *required* to present full validity information of their instrument. That is, if journal editors would insist that authors show how and why their instruments are valid, some unnecessary aspects of such debates might be avoided. Controversies over instrumentation are not uncommon in the field (Bostrom, et al., 1990; Bostrom, et al., 1993; Ruble and Stout, 1993).

5. CONCLUSION

Some real progress has been made in validating IS research that is both quantitative and positivist over the last eleven years. The proportions of empirical studies that verify their instruments through reliability, pre-tests and pilot tests, and the other validities are all elevated. But the field still has ground to make up to reach more comfortable levels of validation. In no single area other than reliability have we reached a point where those researchers using validation techniques outnumber those who do not. Clearly, some of our journals are moving ahead on this front with more vigor than others, but the problem is likely not confined to certain journals or classes of journals.

As a field, we should seek for full validation. The best social science methodologists (Cook and Campbell, 1979; Cronbach, 1971, 1990; Stone, 1979) allow a great deal of latitude in choice of method, but make it clear that the ideal study will use the variety of pretesting and statistical tools to provide evidence that the instrumentation is reliable and valid. Across-the-

board validation of our research, regardless of the methodological approach, could be our next major community goal. Further heuristics and guidelines for bringing even more rigor to the process of positivist, quantitative research need to be proffered. Therefore, for the moment, it is critical that the gatekeepers of the field, as represented by the journals, raise the level of awareness of the entire community by insisting on the four guidelines offered here or ones similar to these.

There are several areas that call for future research on IS validation practices.¹⁴ The quantitative, positivist empirical research reviewed in this paper described trends and aided in the measurement of goal accomplishment. But there is other empirical research that would be highly useful in its own right. Studies of the interaction between various research methodologies and the maturity of scientific fields are valuable in determining the nature of the research endeavor and benchmarking the progress of specific fields. Moreover, empirical work can contribute to the dissemination of knowledge in research communities, which has studied the relationship between paradigms and strong emphases on rigor (Kuhn, 1970; Webster and Starbuck, 1988). One viable approach might be citation analysis to understand these underlying mechanisms, as in the studies by Culnan and associates (Culnan, 1986; Culnan, 1987; Culnan, et al., 1990; Culnan and Swanson, 1986). Of course, all such empirical research would necessarily need to involve rigorous theory, rather than simply involve fishing expeditions in a sea of data.

Specific recommendations for which practices are mandatory and which are simply highly desirable for various kinds of research is badly needed to give guidance and direction to the generations of new IS researchers now entering the field.

¹⁴ Researchers interested in publishing instrument development work may want to review the *MIS Quarterly's* official editorial policy on instrument-development submissions, which appeared in the September 1999 issue. This policy is excerpted in the Appendix. Validation papers on new constructs in the IS field will not be publishable in *MISQ*, but are being encouraged in other venues, such as the *Information Systems Research* Special Issue on e-

Finally, although this paper has studied what is termed “quantitative, positivist” research rather than “qualitative, post-positivist” research, validation principles are also necessary within post-positivist paradigms. Many scholars in the qualitative IS community believe that there are striking differences between the validation principles in these two approaches to research (Leininger, 1994; Denzin and Lincoln, 1994; Lacity and Janson, 1994). Research efforts to provide guiding principles or evaluation criteria to post-positivist researchers have started to emerge.¹⁵ In our opinion, such guidance is as important and necessary as it is in positivist research.

The next decade could see heightened validation standards in all IS research. If so, the IS discipline as a whole will greatly profit from increased research rigor.

Commerce Metrics (<http://isr.commerce.ubc.ca/call.html>). Editorial policies will vary from journal to journal and IS researchers need to be cognizant of these preferences and/or philosophical stances.

¹⁵ An example of such an effort is the Special Issue on Intensive Research being published serially by *MIS Quarterly*. This special issue has already produced four articles in the past year and a half (i.e., Klein and Myers, 1999; Walsham and Sahay, 1999; Trauth and Jessup, 2000; and Schultze, 2000).

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.
- 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.
- Bostrom, R.P., Olfman, L. and Sein, M.K. "The Importance of Learning Style in End-User Training," *MIS Quarterly* (14:1, March), 1990, pp. 101-119.
- Bostrom, R.P., Olfman, L. and Sein, M.K. "Learning Styles and End-User Training: A First Step," *MIS Quarterly* (17:1 (March), 1993, pp. 118-120.
- 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.
- Culnan, M.J. "The Intellectual Development of Management Information Systems, 1972-1982: A Co-Citation Analysis," *Management Science*, (32:2, February), 1986, pp. 156-172.

Culnan, M.J. "Mapping the Intellectual Structure of MIS, 1980-1985: A Co-Citation Analysis," *MIS Quarterly*, September, 1987, pp. 341-353.

Culnan, M.J., O'Reilly, C.A., III and Chatman, J.A. "Intellectual Structure of Research in Organizational Behavior, 1972-1984: A Cocitation Analysis," *Journal of the American Society for Information Science*, (41:6, September), 1990, pp. 453-458.

Culnan, M.J. and Swanson, E.B. "Research in Management Information Systems, 1980-1984: Points of Work and Reference," *MIS Quarterly*, September, 1986, pp. 289-302.

Denzin, N.K. and Lincoln, Y.S. "Introduction: Entering the Field of Qualitative Research," In *Handbook of Qualitative Research*, N.K. Denzin and Y.S. Lincoln (Ed.), Sage, London, 1994, pp. 1-18.

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

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.

Hardgrave, B. and Walstrom, K. "Forums for MIS Scholars," *Communications of the ACM* (40:11), 1997, pp. 119-124.

Holsapple, C.W., Johnson, L.E., Manakyan, H. and Tanner, J. "Business Computing Research Journals: A Normalized Citation Analysis," *Journal of Management Information Systems* (11:1), 1994, pp. 131-140.

Jenkins, A.M. "Research Methodologies and MIS Research," In *Research Methods In Information Systems*, E. Mumford, R. A. Hirschheim, G. Fitzgerald and A. T. Wood-Harper (Ed.), North Holland, Amsterdam, 1985, pp. 103-117.

Klein, H.K. and Myers, M.D. "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems," *MIS Quarterly* (23:1), 1999, pp. 67-93.

Kuhn, T.S. *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, IL, 1970.

Lacity, M.C. and Janson, M.A. "Understanding Qualitative Data: A Framework of Text Analysis Methods," *Journal of Management Information Systems* (11:2, Fall), 1994, pp. 137-155.

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.

- Leininger, M. "Evaluation Criteria and Critique of Qualitative Research Studies," In *Critical Issues in Qualitative Research Methods*, J. M. Morse (Ed.), Sage Publications, Thousand Oaks, 1994, pp. 95-115.
- 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.
- Nord, J. H. and Nord, G. D. "MIS Research: Journal Status Assessment and Analysis," *Information & Management*, 29, 1, (1995), 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.
- Pitt, L.F., Watson, R.T. and Kavan, C.B. "Service Quality: A Measure of Information Systems Effectiveness," *MIS Quarterly* (19:2, June), 1995, pp. 173-187.
- Pitt, L.F., Watson, R.T. and Kavan, C.B. "Measuring Information Systems Service Quality: Concerns for a Complete Canvas," *MIS Quarterly* (21:2, June), 1997, pp. 209-221.
- Rogers, T.B. *The Psychological Testing Enterprise*, Brooks/Cole Publishing Company, Pacific Grove, CA, 1995.
- Ruble, T.L. and Stout, D.E. "Learning Styles and End-User Training: An Unwarranted Leap of Faith," *MIS Quarterly* (17:1, March), 1993, pp. 115-118.
- Schultze, U. "A Confessional Account of an Ethnography about Knowledge Work," *MIS Quarterly* (24:1), 2000, pp. 3-41.
- Schwandt, T.A. *Qualitative Inquiry: A Dictionary of Terms*, Sage Publications, Thousand Oaks, CA, 1997.
- 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.
- 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.
- Sussmann, M. and Robertson, D.U. "The Validity of Validity: An Analysis of Validation Study Designs," *Journal of Applied Psychology* (71:3), 1986, pp. 461-468.
- Trauth, E.M. and Jessup, L.M. "Understanding Computer-Mediated Discussions: Positivist and Interpretive Analysis of Group Support System Use," *MIS Quarterly* (24:1), 2000, pp. 43-79.
- 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.
- Van Dyke, T.P., Kappelman, L.A. and Prybutok, V.R. "Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire," *MIS Quarterly* (21:2, June), 1997, pp. 195-208.
- Walsham, G. and Sahay, S. "GIS for District-Level Administration in India: Problems and Opportunities," *MIS Quarterly* (23:1), 1999, pp. 39-65.
- Walstrom, K.A., Hardgrave, B.C. and Wilson, R.L. "Forums for Management Information Systems Scholars," *Communications of the ACM* (38:3), 1995, pp. 93-106.
- Webster, J. "Desktop Videoconferencing: Experiences of Complete Users, Wary Users, and Non-Users," *MIS Quarterly* (22:3), 1998, pp. 257-286.
- Webster, J. and Starbuck, W.H. "Theory Building in Industrial and Organizational Psychology," In *International Review of Industrial and Organizational Psychology*, John Wiley & Sons, Ltd., London, 1988, pp. 93-138.
- Yin, R.K. *Case Study Research: Design and Methods*, Sage Publications, Thousand Oaks, CA, 1994.

APPENDIX

The following is excerpted from the “Editor’s Comments” in the September, 1999 issue of *MISQ*:

“Clarification in Editorial Policy on Instrument-Development Submissions

The senior editors and I have decided that *MIS Quarterly* will no longer accept submissions that are primarily about the development, reliability, or validity of instruments. This includes manuscripts on instrument-development work where the objective is primarily about developing a valid and reliable instrument for a construct that currently is not firmly grounded in existing theory used in information systems research.

Not all manuscripts involving instrument development are being ruled out. The following are overlapping categories in which an appropriate submission, involving instrument development, could fall.

1. The manuscript, although mainly oriented towards measurement issues, also does a good job of developing a construct. In making a significant contribution to theory in this way, the manuscript could be an appropriate submission.
2. The manuscript describes work that seeks to develop a valid and reliable instrument for a construct that is central to a well-established theory used in information systems research and for which no valid and reliable instrument currently exists.
3. The manuscript describes work that shows how to improve, quite substantively, the validity and reliability of measures of a construct that is well established in the information systems field. The focus here is on substantive improvements over existing instruments. The authors might be proposing a new instrument that is demonstrably better than existing instruments or modifications to existing instruments that will improve their validity and reliability considerably.
4. The authors may be building a new theory about some type of information systems phenomena, and their instrument-development work forms part of their efforts to test the theory. Thus, the instrument-development work is simply part of the larger work to build and test theory.

MIS Quarterly also discourages and does not participate in the practice in which authors present the development of an instrument in one manuscript and then submit another manuscript that uses the instrument to test a research model with the same data set.”