Validation in information systems research: A state-of-the-art assessment
Marie-Claude Boudreau; David Gefen; Detmar W Straub
MIS Quarterly; Mar 2001; 25, 1; ABI/INFORM Global
pg. 1

By: Marie-Claude Boudreau
Management Information Systems
Department
Terry College of Business
University of Georgia
Athens, GA 30602
U.S.A.
mboudre@terry.uga.edu

David Gefen
Management Department
LeBow College of Business
Drexel University
Philadelphia, PA 19104
U.S.A.
gefend@drexel.edu

Detmar W. Straub
Department of Computer Information
Systems
College of Business Administration
Georgia State University
Atlanta, GA 30302-4015
U.S.A.
dstraub@gsu.edu

Abstract

Over 10 years ago, the issue of whether IS researchers were rigorously validating their quantitative, positivist instruments was raised (Straub 1989). 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 a recent three year period for use of validation techniques. Findings suggest that the field has advanced in many areas, but, overall, it appears that a majority of published studies 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.

1Allen Lee was the accepting senior editor for this paper.
Keywords: IS research methods, measurement, psychometrics, validation, reliability, content validity, construct validity, manipulation validity, quantitative research, positivist research, guidelines.

ISRL Categories: AI0401, AI0301, AI0402, AI0403, AI801

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

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 1960s. 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 straightforward 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 items 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.

Method

In order to determine whether instruments used by the IS community are more thoroughly validated now than 11 years ago, a new review and analysis of the literature was conducted. By maintaining a similar spirit to Straub’s (1989) research, comparability between the two time periods is possible. We do not pretend, however, to directly replicate Straub’s work. Adjustment for journal selection and sampling and coding procedures, as will be discussed next, were indeed necessary.

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 1980s. Because the period of inquiry for the current research covers the three years between 1997 and 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.footnote[2] Among these nine journals, Nord

footnote[2]In alphabetical order, these journals were Communications of the ACM, Decision Sciences, Information & Management, Information Systems Management, Jour-
and Nord concluded that six published "important" or "very important" research. 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 that 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. 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 1999. As in Straub, the qualifying criteria for the sample was that the article employed either (1) correlational or statistical manipulation of variables or (2) 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.

To enrich the analysis, we expanded the attributes collected by Straub. In total, 11 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. 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. We felt it was important to assess type of research 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 prior studies (Alavi and Carlson 1992; Alavi et al. 1989; Stone 1979), 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 (1978, 1979) 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-

---

3 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.
gathering techniques, field studies can employ either questionnaires, administered in person, by mail or e-mail, 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 et al. (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.

With these four types of research methods, we classified each sampled article. One reason it was desirable to capture research method type was that, in 1989, 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 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

intense inquiry of a single site, and because not all of 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.

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).

Benbasat et al. (1987) also mentioned, 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 (1995) research. Webster investigated the use of videoconferencing in a 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

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, the findings of this study will tend to overestimate the extent to which the field is pursuing appropriate validation efforts.
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 pretests and pilot tests have 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 this validity. Consistent with Straub, 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 (Straub 1989; Zaltman et al. 1973). 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.6

Establishment of the **reliability** of an instrument, or the absence of the same, was also tallied. As pointed out in Rogers, 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. 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 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 11 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, 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 and, as such, there can be no assurance of the internal validity of the experiment without a verification that the manipulations have "taken" (Perdue and Summers 1986). 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

---

6It should be noted that some conceptualize predictive validity as separate and distinct from construct validity (e.g., Bagozzi 1980; Campbell 1980; Cronbach 1990). 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).
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 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 and we felt that it was important to see if IS researchers, 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 in 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.

Study Results

Sample

A total of 193 articles were used in the analysis.⁹

Among the articles reviewed, 24 originated in MIS Quarterly, 26 in Information Systems Research, 81 in Information & Management, 50 in Journal of Management Information Systems and 12 in Management Science.¹⁰ Whereas most were field studies (64%), coded works also included laboratory experiments (25%), case studies (6%), and field experiments (5%).

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 11 coded attributes, the following percentages of agreement were obtained: type of research—74%; research method—81%; pre-test—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 et al. (1989). For all criteria, the average kappa was 0.75, which is above the 0.70 inter-rater minimum reliability recommended by Bowers and Courtright (1984), Landis and Koch (1977), and Miles and Huberman (1994). Before further analysis was performed on the collected data, disagreements between the coders were reconciled by one of the authors.

---

⁹Overall, 394 articles were considered for inclusion. More particularly, the distribution of articles was as follows: 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 201 articles, overall.

¹⁰The overwhelming number of selected articles from Information & Management can be explained by its publication frequency, which is monthly rather than quarterly.
Overview of Findings

As indicated in Table 1, the data clearly shows that instruments are more frequently validated than they were 11 years ago. Nevertheless, it appears that the proportion of published studies validating their instruments is smaller than the proportion of published studies not validating their instruments in all categories, except for the use of reliability. Both pretest and pilot studies were more frequently employed, that is, 47% 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 23% and 37% in the current survey. With regard to reliability, these tests are used more widely. Whereas 17% of the surveyed studies assessed reliability in 1989, 63% perform these tests now.

On this matter, it is worth mentioning that the majority of the studies assessing reliability of their instruments have done so through the standard coefficient of internal consistency, i.e., Cronbach's $\alpha$ (79%). Only in rare cases have other methods been used to verify reliability of measures. Specifically, 2% used test/retest, 2% used split halves, and 21% 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, such as 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 that of the other journals, but sufficient for comparison purposes, so there is no ready explanation for this deficiency.

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 not using one (44% vs. 32%; 74% vs. 54%). 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 (50% vs. 43%; 25% vs. 20%).

Another telling point is how confirmatory studies (62 articles or 32% of total) compare to exploratory studies (131 articles or 68% of total). The present survey indicates that, as we suspected, exploratory studies showed less interest in validating their instruments than confirmatory studies (refer to 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.

Another way to look at this data is to consider the type of research and the nature of an instrument criteria together. On the one hand, the sampled confirmatory studies used an existing instrument almost as often as they created a new instrument (48% vs. 52%). On the other hand, the sampled exploratory studies developed a new instrument more frequently (60%) than they used an existing instrument (40%). 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.

---

11The 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 published studies assessing each of the different validation criteria were approximately the same over the years 1997, 1996, and 1999.
### Table 1. Survey of Instrument Validation Use in MIS Literature

<table>
<thead>
<tr>
<th>Instrument Categories</th>
<th>2000</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I&amp;M</td>
<td>ISR</td>
</tr>
<tr>
<td>Pretest</td>
<td>27%</td>
<td>35%</td>
</tr>
<tr>
<td>Pilot</td>
<td>32%</td>
<td>19%</td>
</tr>
<tr>
<td>Pretest or Pilot*</td>
<td>47%</td>
<td>46%</td>
</tr>
<tr>
<td>Previous Instrument</td>
<td>35%</td>
<td>58%</td>
</tr>
<tr>
<td>Utilized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Validity</td>
<td>16%</td>
<td>31%</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>31%</td>
<td>54%</td>
</tr>
<tr>
<td>Reliability</td>
<td>56%</td>
<td>85%</td>
</tr>
</tbody>
</table>

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

### Table 2. Studies with Previously Utilized Instrument versus Those with New Instrument

<table>
<thead>
<tr>
<th>Instrument Categories</th>
<th>Previous Instrument (N = 82)</th>
<th>New Instrument (N = 111)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tot N</td>
<td>Tot %</td>
</tr>
<tr>
<td>Pretest or Pilot</td>
<td>35</td>
<td>43%</td>
</tr>
<tr>
<td>Content Validity</td>
<td>16</td>
<td>20%</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>36</td>
<td>44%</td>
</tr>
<tr>
<td>Reliability</td>
<td>61</td>
<td>74%</td>
</tr>
</tbody>
</table>

### Table 3. Type of Research (Confirmatory versus Exploratory Studies)

<table>
<thead>
<tr>
<th>Instrument Categories</th>
<th>Confirmatory Studies (N = 62)</th>
<th>Exploratory Studies (N = 131)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tot N</td>
<td>Tot %</td>
</tr>
<tr>
<td>Pretest or Pilot</td>
<td>29</td>
<td>47%</td>
</tr>
<tr>
<td>Content Validity</td>
<td>22</td>
<td>35%</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>33</td>
<td>53%</td>
</tr>
<tr>
<td>Reliability</td>
<td>43</td>
<td>69%</td>
</tr>
</tbody>
</table>

### Table 4. Relationship Between Type of Research and Nature of an Instrument

<table>
<thead>
<tr>
<th>Nature of an Instrument</th>
<th>Confirmatory Studies (N = 62)</th>
<th>Exploratory Studies (N = 131)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tot N</td>
<td>Tot %</td>
</tr>
<tr>
<td>Previous Instrument</td>
<td>30</td>
<td>48%</td>
</tr>
<tr>
<td>New Instrument</td>
<td>32</td>
<td>52%</td>
</tr>
</tbody>
</table>
A 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, all of the previously introduced validity criteria were proportionally less prominent in laboratory and field experiments than they were in field studies. However, when comparing case studies to field studies, the story is slightly different. Although published field studies are more inclined to assess content and construct validity and to use a pretest or pilot test to validate their instrument, they are less likely to assess reliability than are case studies.

Straub encouraged researchers to include an "Instrument Validation" subsection when accounting for results. The present research found that only 24% of the surveyed articles did so over the investigated period. Interestingly enough though, this percentage was not uniform across the surveyed journals. Whereas more than 42% of selected articles in Information Systems Research had this specific subsection, less than 30% 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, 29%, 25%, 20%, and 19% 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 such a subsection, whereas only 6% of the articles in Management Science did so.
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 checks 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, reviewers may not require these checks even though the researchers had actually performed them. 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 them.

Do studies using first generation statistical techniques (regression, ANOVA, LOGIT, etc.) differ from studies using second generation statistical techniques (e.g., Structural Equation Modeling [SEM] such as LISREL and PLS) 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. Indeed, there is a difference of 53% in the proportions of studies assessing construct validity if we compare SEM to non-SEM studies. Similarly, there is a difference of 39% 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 hypothesis testing and, thus, do not determine the overall fit indexes.

### Summary of Key Points

Overall, despite the improvements made over the previous 11 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:

---

12The 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.
1. Published studies 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 $\alpha$; alternative methods, or a combination of methods, for validation of mature research streams are rarely used.

3. Published studies 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 experiments, as well as case studies, lag behind field studies with respect to most validation criteria.

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. 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.

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.

Discussion

Overall, the data clearly shows that there has been 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 to make more rigorous and credible the instruments used to conduct quantitative, positivist IS 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 quantitative, positivist IS studies are validating their instruments today, but the field has still not reached the point where validation is the rule rather than the exception. In only one category in our investigation did the number of published studies engaging in a validation practice exceed those that 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.
One point in the 1989 guidelines calls for further clarification based on practice regarding this specification since the publication of the 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 revalidating 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 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. (1989, 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). This practice, thus, would be highly desirable for research that is positivist and quantitative.

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 upfront in the article itself.

A good example of this situation is the SERVQUAL instrument, which was introduced into IS research by Pitt et al. (1995) and was later challenged by Van Dyke et al. (1997). Whereas Pitt et al. (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 requested to present full validity information of their instrument in the first place. 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, 1993; Ruble and Stout 1993).
Conclusion

Some real progress has been made in validating IS research that is both quantitative and positivist over the past 11 years. The proportions of empirical studies that verify their instruments through reliability, pretests and pilot tests, and the other validities are all increased. 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 published studies using validation techniques outnumber those that 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 must strive for validity in all of our research. The best social science methodologists (Cook and Campbell 1979; Cronbach 1971, 1990; Stone 1978) 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 proposed. 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. The qua-

titative, 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 that has studied the relationship between paradigms and rigor can contribute to the dissemination of knowledge in research communities (Kuhn 1970; Webster and Starbuck 1988). One viable approach to understand these underlying mechanisms might be citation analysis, as in the studies by Culnan and associates (Culnan 1986, 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 are badly needed to give guidance and direction to the generations of new IS researchers now entering the field.

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 (Denzin and Lincoln 1994; Lacity and Janson 1994; Leininger 1994). Research efforts to provide guiding principles or evaluation criteria to post-positivist researchers have started to emerge. In our

---

Guidelines for the appropriate use of first and second generation statistical techniques in positivist research may be found in Gefen et al. (2000).

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 and is also available at http://www.misq.org/archivist/volno23/issue3/edstat.html#instrument. It explains that validation papers on new constructs in the IS field will be publishable in MISQ, but only under certain conditions. Validation papers on new constructs are being encouraged in other venues, such as the Information Systems Research Special Issue on e-Commerce Metrics. Editorial policies will vary from journal to journal and IS researchers need to be cognizant of these preferences and/or philosophical stances.

An example of such an effort is: the Special Issue on Intensive Research published serially by MIS Quarterly. This special issue produced six articles in the past year and a half (i.e., Gopal and Prasad 2000; Klein and Myers 1999; Nelson et al. 2000; Schultz et al. 2000; Truth and Jessup 2000; Walsham and Sahay 1999).

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
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.

References


About the Authors

Marie-Claude Boudreau is an assistant professor of MIS at the University of Georgia. She received a Ph.D. degree in CIS from Georgia State University, a Diplôme d’Enseignement Supérieur Spécialisé from l’École Supérieure des Affaires de Grenoble, and an M.B.A. from l’Université Laval in Québec. Her current research investigates the consequences of information systems in organizations. She has published in Information Systems Research, The Academy of Management Executive, Information Technology & People, and Communications of the AIS.

David Gefen is an assistant professor of MIS at Drexel University, where he teaches Strategic Management of IT, Database Analysis and Design, and Programming languages at the MBA level. He received his Ph.D. degree in CIS from Georgia State University and a Master of Sciences from Tel-Aviv University. David has conducted extensive research on issues relating to the adoption, implementation and use of advanced computer technologies. His wide interests in IT adoption stem from his 12 years of experience in developing and managing large information systems, including the ongoing management of a large state-of-the-art logistics system. His research specialization is in IT adoption, the Internet, culture and gender effects, and e-trust. His current research interests focus on psychological and relational processes involved in the successful implementation of technological innovations. His research findings have been published or are forthcoming in leading academic and professional journals, including the MIS Quarterly, DATA BASE, Omega: The International Journal of Management Science, Journal of the AIS, eServices Quarterly, and Communications of the AIS. David is also the author of several encyclopedia articles on IT adoption and gender, e-Commerce, and IT security, and a book chapter on trust and ERP adoption.

Detmar W. Straub is the J. Mack Robinson Distinguished Professor of Information Systems at Georgia State University. He has conducted research in the areas of e-Commerce, computer security, technological innovation, international IT studies, and IS research methodologies. He holds a DBA in MIS from Indiana and a Ph.D. in English from Penn State. He has published over 85 papers in journals such as Management Science, MIS Quarterly, Information Systems Research, Organization Science, Communications of the ACM, Journal of MIS, Journal of Global Information Management, Information & Management, Communications of the AIS, Academy of Management Executive, and Sloan Management Review. He is currently an Associate Editor for Management Science and Information Systems Research. Former Co-Editor of DATA BASE for Advances in Information Systems and an Associate Editor and Associate Publisher for MIS Quarterly, he has consulted widely in industry in the computer security area as well as in the areas of e-Commerce and technological innovation. He teaches courses at Georgia State in the areas of: Electronic Commerce Strategy, IT Strategies for Management, Systems Integration and IT Outsourcing, International IT Policies and Issues, and Computer Security Management.

16 MIS Quarterly Vol. 25 No. 1/March 2001