ELSEVIER

Journal of Operations Management 16 (1998) 407–425

JOURNAL OF OPERATIONS MANAGEMENT

An assessment of survey research in POM: from constructs to theory

Manoj K. Malhotra *, Varun Grover [1]

Management Science Department, University of South Carolina, Columbia, SC 29208, USA

Abstract

The use of field-based empirical methodologies in the production and operations management (POM) area has been steadily increasing over the past several years. One of the most prominent among these is the survey research methodology which has often been used to capture data from business organizations. However, to effectively contribute to theory development in the POM field, this methodology must be carefully implemented. Poorly designed and executed survey research is of little or no value. This paper attempts to provide a normative perspective on 'good survey research practices'. In doing so, it attempts to bridge the gap between survey research and theory development. The paper liberally draws from work in other social science disciplines like psychology, marketing, organizational behavior, and other related fields that are more mature in the use of surveys, and applies it to the POM area. A set of ideal survey research attributes are identified, and existing POM survey work in four prestigious academic journals is evaluated with respect to these attributes. Results of assessing 25 survey-based articles published in these journals between 1990 and 1995 indicate that both exploratory as well as explanatory survey research is being conducted, with current emphasis being more on explanatory kind of research. In order to move the field forward, greater attention will need to be focused on employing multi-item constructs, assessing them for content validity, and purifying them through field-based pretesting. More sophisticated POM studies in the future would have to aim for testing theory with reliable and valid scales that are relatively free of measurement related errors. It is hoped that this work will act as a catalyst in compressing the learning curve with respect to survey research practices in POM, and also accelerate the use of greater methodological rigor in future empirical studies in the field.

Keywords: Operations strategy; Survey research methods; Research validity; Construct measurement

1. Introduction

There is increasing pressure on business schools to produce 'relevant' research. While the notion of relevance is debatable, it has triggered a recognition of the importance of field-based research in which data is gathered from the business context or the social setting in which practice occurs. This is particularly true in the production and operations management (POM) field, which has witnessed an increasing number of 'nontraditional' field-based methodologies in recent years as opposed to the more 'traditional' ones involving simulation and modeling. While diversity of methods is generally positive and indicates the maturing of a field, it is critical that the methods be appropriately implemented in order to

* Corresponding author. Tel.: +1-803-777-2712; fax: +1-803-777-6876; e-mail: malhotra@darla.badm.sc.edu

[1] Tel.: +1-803-777-5923; fax: +1-803-777-6876; e-mail: vgrover@darla.badm.sc.edu

0272-6963/98/$19.00
PII S0272-6963(98)00021-7

obtain meaningful and valid results. Inability or unwillingness to do so could impede the production of knowledge or theory which is very dependent on techniques for collecting, analyzing, and interpreting data (Simon, 1980).

Survey research is prominent as a methodology that has been used to study unstructured organizational problems in the POM area. While the use of survey research is relatively recent in POM, the basic methods have been known since the 1950s and have continued to evolve since that period. Often, however, studies fail to adhere to fundamental guidelines in their conduct of survey research. In many cases, these guidelines are implicit in books and articles on methodology, but may not be well articulated as standards for the field. In other cases, individuals in their expediency to get published overlook normative standards. This can be disastrous for the field for essentially one far reaching but simple reason: no single finding in such a study can be trusted!

On a broad level, any research should be carried out in a systematic and programmatic manner (Hunter et al., 1983; Flynn et al., 1990). For instance, greater attention to (say) instrumentation in survey research promotes cooperative efforts and permits confirmatory follow-up research to use a tested instrument. Validated instruments of both independent and dependent variables can alleviate confounding effects in determining the true relationship among variables. Also, careful pretesting of instruments in the field can serve as a reality check indicating to the researcher how well conceptualizations of the problem match the actual experience of the practitioner. Careful adherence to such standards can prevent survey research studies from becoming isolated and the field from becoming fragmented, i.e., without a cumulative tradition of building on prior work that is so essential for knowledge building.

This study addresses issues related to survey research in the POM area. Flynn et al. (1990) provided an excellent overview of general methodology that should be followed for conducting empirical research in the field. This overview included establishing the theoretical foundation for the research, selecting an appropriate research design and data collection method, properly implementing the study, and finally using correct data analysis techniques for interpreting the results. These authors indicated that survey designs with questionnaires are the most commonly used methodology in empirical POM research. However, it is not clear as to what extent good research practices have been subsequently followed by POM scholars in conducting their own survey-based research. To perform such an evaluation, an exhaustive elicitation of attributes that characterize high quality survey research is needed. This information while available in different books and journals, is usually not cohesive or interwoven into an integrated framework that can facilitate prescription. Its consolidation here would be helpful to future POM researchers engaging in survey-based work, and perhaps also lead them to avoid pitfalls that have commonly occurred in other social science disciplines that pioneered the use of this methodology. For example, it has been shown in a mature field like marketing that large measurement related errors exist in defining important constructs. Cote and Buckley (1987) in their meta-analysis of 64 studies reported that trait (or actual construct related) variance accounts for less than 50% of the observed variance in measures, with more abstract constructs being considerably more difficult to measure. This compromises the reliability of these studies, as also noted by the authors—''given the demonstrably poor quality of some of the measures that have been used in consumer behavior research, researchers should be cautious about evaluating or comparing alternative theories based solely upon empirical evidence unless the appropriateness (validity) of some of the measures has been determined'' (Cote and Buckley, 1988).

These observations have major implications for theory building in the POM field. It is clear that we cannot build upon prior work without evaluating the methodological rigor of existing survey-based research in POM. To do this assessment, we first develop an exhaustive list of criteria from the knowledge based on other mature social science disciplines. While we do not directly evaluate here the quality of constructs that have been used in prior POM research, it is important to understand how construct measurement-related errors get introduced into survey-based studies, and what can be done to minimize them. Accordingly, the following questions are addressed:

1. What is survey research and what is its relationship to theory development?

2. Based on this relationship, what are the 'norms' (ideal attributes) of conduct for good survey research?
3. How can current survey research in POM be evaluated with respect to these norms, and where are the avenues for improvement?

Collectively, these questions address normative, descriptive, and prescriptive aspects of survey research as it relates to the POM field. The paper is also generally organized by sections in the order of the three questions listed above. However, we first establish the various modalities of survey research at the outset in an order to better demarcate the domain of this study. Subsequently, we present implications of our findings and the related conclusions for future research directions.

2. The 'what' and 'when' of survey research

In general, a survey involves the collection of information from a large group of people or a population. Within the bounds of this definition are a variety of data elicitation processes including opinion surveys, political polls, TV viewing polls, etc. This paper excludes such surveys. In contrast, the focus is on *survey research*, which is conducted to advance scientific knowledge or develop theory.

Survey research has three distinct characteristics. First, it involves collection of information by *asking people* for information in some structured format. Depending on the quality and cost tradeoffs involved, collection of information or data could take place using mail questionnaire, telephone interview, or face-to-face interview. Depending on the unit of analysis, the individuals surveyed could be representatives of themselves, their project, their expertise, or their organization. Second, survey research is usually a *quantitative method* that requires standardized information in order to define or describe variables, or to study relationships between variables. Third, information is gathered *via a sample*, which is a fraction of the population, with the need to be able to generalize findings from the sample to the population. Given these three characteristics, it is easier to distinguish survey research from other field-based methods like case studies. Case studies are not usually quantitatively oriented, the variables are often not predefined, and such studies involve examination of a phenomenon in depth within their natural setting, thereby precluding any attempt at generalization.

There are two major types of survey research (Kerlinger, 1986). The first type can be classified as 'exploratory' and the objective is to become more familiar with a topic. There is usually no model in exploratory research and the concepts of interest need to be better understood and measured. An exploratory survey is useful in determining, for instance, the benefits that may be associated with adopting MRP systems and problems that impede its successful implementation. Resulting data can then be refined to identify new possibilities and dimensions of interest. Another type of exploratory survey research is referred to as 'descriptive'. This type of study has been described as indispensable in the early stages of studying a phenomenon (Dubin, 1978) as it develops the units that comprise theories. In other words, descriptive research is aimed at describing the distribution of a phenomena in a population, thereby ascertaining facts. For example, a descriptive survey might be concerned with documenting the types of manufacturing processes being used by small and large manufacturing firms. Hypothesis related to common perceptions or changes over time could be formulated and tested.

The second type of survey research and arguably the most important is 'explanatory research'. This research is devoted to finding causal relationships among variables. It does so from theory-based expectations on how and why variables should be related. Hypotheses could be basic (i.e., relationships exist) or could be directional (i.e., positive or negative). For instance, an explanatory study could explain, hypothesize, and test for a positive relationship between the existence of an MRP system and success in materials management. Results then are interpreted and in turn contribute to theory development.

Drawing causal inferences (i.e., a change in variable A causes a change in variable B) is very difficult given survey research designs. The most common design is *cross-sectional* in which information is collected at one point in time from a sample chosen to represent the population. The lack of a temporal dimension makes it difficult to establish causality, but the design is appropriate to test differences in population subsets. In contrast, *longitudinal*

designs are appropriate for studying phenomenon that change (e.g., the impact of MRP implementation on work design over time) by collecting data in the same organization at two or more points over time. These designs are difficult to implement, but can provide greater confidence regarding causality. Regardless of whether a cross-sectional or longitudinal design is used, it is imperative that the *unit of analysis* be clearly defined at the outset. In other words, all questions in the instrument should be collecting information at a consistent unit of analysis, whether it be the individual, work group, project, function, organization or even industry.

It should be emphasized that the ultimate aim of survey research is to contribute to theory development. Alternatively stated, survey research should better explain or predict a phenomena after evolving through the maturity cycle as shown in Fig. 1. To do so, consistent relationships between the various theoretical concepts must be established and verified through continuous testing and extension. As can be seen in Fig. 1, evolution of research into a phenomena increases understanding or the certainty with respect to knowledge. Exploratory and descriptive surveys can help identify the concepts and the basis for measurement, and are very appropriate for early stages of the research. As the research matures, variables can be effectively measured and relationships can be studied using explanatory surveys. In many cases *triangulation* or the use of multiple methods (both qualitative and quantitative) to cross-check each other, is desirable and can enhance confidence in findings.

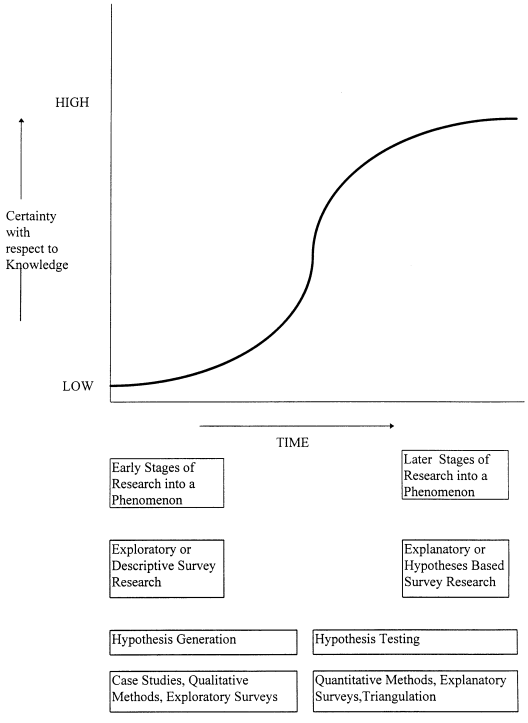

Fig. 1. The maturity cycle of research.

3. Survey research and theory development

Kerlinger (1986) defines theory as: ''a set of interrelated constructs (concepts), definitions and propositions that present a systematic view of phenomena by specifying relationships among variables, with the purpose of explaining and predicting the phenomena''.

This definition spans two domains. One can be labeled the *theoretical domain* and the other the *operational domain*. Constructs or concepts are abstractions in the theoretical domain that express similar characteristics (e.g., intelligence, organizational success, manufacturing effectiveness). These constructs are 'latent' or are not directly observable or measurable (Bagozzi, 1979). Therefore, theory attempts to explain observed phenomena by systematically setting out interrelationships between constructs. However, since these constructs are latent, researchers must provide an operational definition of it that is observable. This operational definition represents a symbol or a variable to which numeric values can be assigned.

Fig. 2 illustrates the relationship between the theoretical domain and the operational domain. The theoretical domain expresses a relationship of interest between latent constructs 'X' and 'Y'. The operational domain examines a corresponding relationship between the variables '*x*' and '*y*' in the operational domain. Theory (or the construct relationships in the theoretical domain) is invaluable in classical confirmatory research (deductive research) since it prespecifies the makeup and structure of the constructs and can guide propositions or hypothesis to be tested in the operational domain. The results of these tests can confirm or modify theory. Robust theory or theory that can withstand scrutiny in multi-

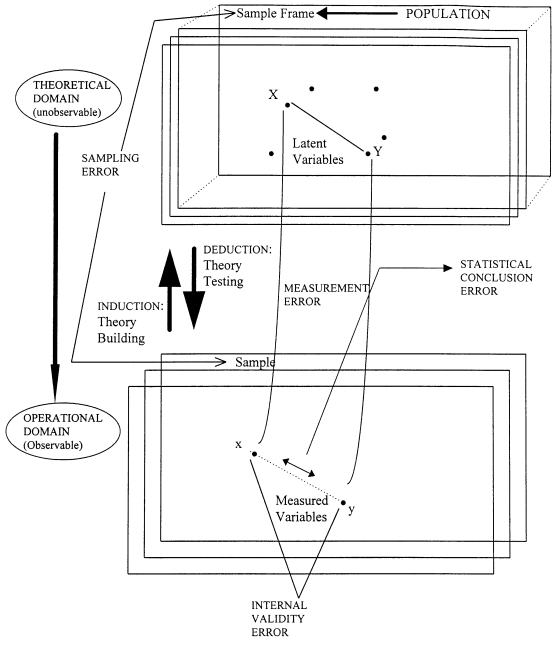

Fig. 2. The link and sources of error between the theoretical and operational domain in survey research.

ple contexts is a noble goal, but is elusive when dealing with the socio-technical systems found in most real world manufacturing and service environments.

It should also be noted that there is an increasingly vociferous group of researchers that espouse inductive research (or theory building) approaches (Eisenhardt, 1989). These approaches can use exploratory techniques (e.g., exploratory surveys) to search for patterns among variables and therefore build theory. Pure deductive research is involved in theory testing, and any failure to confirm hypothesis would result in rethinking theory independent of the data. Inductive approaches would recommend adjustments in the data to make it work and subsequent modification in theory based on these adjustments. A middle ground involving symbiotic interaction between deductive and inductive approaches, theory building and testing, and exploratory and explanatory research, is probably the best representation of the scientific research cycle (McGrath, 1979; Babbie, 1989).

In translating latent constructs to measurable variables, a number of sources of error can be introduced. Survey research must work toward reducing these errors through careful adherence to norms or standards. These errors can be divided into four components (see Fig. 2). *Measurement Error* is the error in measuring latent constructs (i.e., 'X' to '*x*' and 'Y' to '*y*'). Careful validation of the instrument can reduce measurement error. *Sampling error* is the error introduced in selecting the study population and the representativeness of the sample with respect to the population. *Internal Validity Error* reflects the error introduced if other explanations (rival hypothesis) can explain observed relationships. In other words, does '*x*' lead to '*y*' or are there other variables that can explain the change in '*y*'? Finally, *Statistical Conclusion Error* reflects the probability that the null hypothesis has been correctly rejected and that mathematical relationships between hypothesized variables do exist.

Section 4 elaborates on these errors and identifies standards for the conduct of survey research in POM.

4. The conduct of survey research: ideal survey attributes

Before we discuss the various sources of error in survey research, some general attributes discussed earlier need to be formally stated. The first deals with the *unit of analysis*. While the respondent is usually an individual, the unit that person represents must be clearly articulated at the outset and the instrumentation should consistently reflect that unit. If the individual is appropriate for the research question and is responding for himself (or herself), then there is no problem. However, if the unit is the organization, and the individual chosen is low in the organizational hierarchy, bias might be introduced by having a 'functional worker' respond to organizational level variables (e.g., degree of centralization of major decision making in the organization). In other words, the person(s) most knowledgeable about the construct of interest should be chosen (Huber and Power, 1985). Stated as attribute questions are as follows.

(1) Is the unit of analysis clearly defined for the study?

(2) Does the instrumentation consistently reflect that unit of analysis?

(3) Is the respondent(s) chosen appropriate for the research question?

Another general attribute pertains to the notion of *triangulation* or multiple methods to better assess the variables of interest. For instance, written instrumentation (i.e., a multi-item measure), multiple respondents (more than one response per company), interviews (a series of structured open questions) with organizational representatives, and objective measures (financial data), can be used to assess environmental uncertainty faced by organizations. Clearly, such cross validation of base data enhances confidence in results and is desirable, but it could also come at a tremendous cost in terms of both time and effort and may not be practical in many instances.

(4) Is any form of triangulation used to cross validate results?

4.1. Measurement error

Measurement error represents one of the most significant sources of error in survey research. Inappropriate measurement can be due to a number of factors including poorly worded questions, [2] length of instrument, bias induced by method, etc. While measurement error is almost inevitable, the primary question for POM researchers is the extent to which these errors affect the findings. Fortunately, validation techniques are available to reduce measurement error. A framework for doing this, modified from Churchill (1979) seminal work is illustrated in Fig. 3.

It should be noted that the process outlined in Fig. 3 is only applicable to multi-item measures of a variable. As emphasized by Churchill, ''multi-item measures have much to recommend them. First, individual items usually have considerable uniqueness or specificity in that each item has a low correlation with the attribute (latent construct) being measured and tends to relate to other attributes as well. Second, single items tend to categorize people into a relatively small numbers of groups. Thirdly, individual items have considerable measurement error; they produce unreliable responses . . . '' (Churchill, 1979, p. 66). Multi-item measures, on the other hand, can better specify the construct domain, average out uniqueness of individual items, make

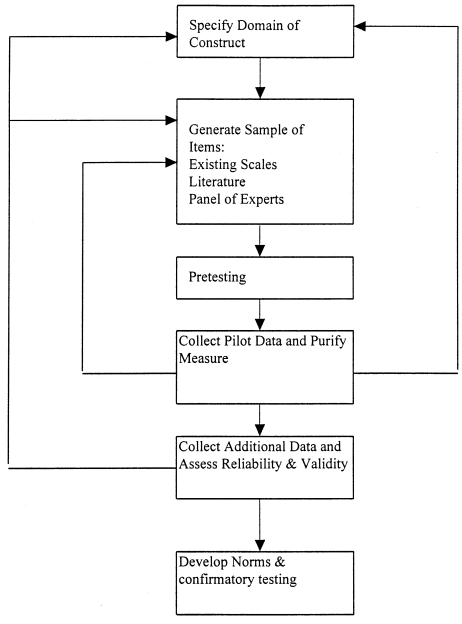

Fig. 3. A framework for developing measurement scales.

fine distinctions between people, and have higher reliability. In developing these measures, however, it is very important that the domain of the construct be well specified and that the items be generated based on this domain. This assessment of the appropriateness of the items to the domain of the construct is referred to as *content validation*. It can be done through the theoretical basis for the items in the literature or a panel of experts who are well versed with the domain. Q-sort techniques can be used where experts are asked to separate items that are relevant to the domain from those that are not. Realistically, however, existing (and preferably validated) scales should be adopted (or adapted) wherever possible in order to cultivate a cumulative tradition of research. Following the development of the preliminary instrument, pretesting with practitioners in the field is highly desirable. This should involve field-based validation of the research, content validity of items, and clarity and wording problems (i.e., biased, ambiguous, inappropriate or double barrel items).

At this stage, actual pilot data should be collected to purify the measure. Internal consistency (or *reliability*) of the items through assessment of Cronbach's Alpha (Nunnally, 1978) should be done to test

[2] Henceforth, a single question will be referred to as an item in this paper.

whether items 'hang together'. [3] Items that do not, can be dropped based on their item/total correlation. Low inter-item correlations can also indicate that the items are not drawn from the same domain. *Test–retest* reliability is another form of reliability that provides evidence of the consistency of the instrument over time. This would require readministration of the instrument and correlational analysis. Other validity testing can also be done. The most common is *construct validation*, which lies at the heart of the scientific process and addresses the question of what the instrument is actually measuring. An internally homogeneous measure might not be measuring what it is supposed to. Two components of construct validity, *convergent* and *discriminant* validity can be assessed. These collectively refer to whether the measure is similar within itself and yet sufficiently different from other measures. Factor analysis using items from multiple measures in the research model has been used to establish construct validity. If the items for each variable load together in factor analysis but do not cross load onto other variables, there is evidence of construct validity. Alternatively, Campbell and Fiske's (Campbell and Fiske, 1959) multitrait multi-method (MTMM) approach can be used where different methods used to measure the same variable should have high correlations relative to correlations across items in different variables. Comparison of correlations in a MTMM matrix can provide indication of construct validity.

Criterion-related validity or predictive validity check must also be performed, and refers to the ability of the scale to predict (or at least relate to) one or more external variables. To do this, however, requires identification of external (criterion) variables that have well documented theoretical relationships with the scale. We chose to drop this attribute here simply because of the difficulty in establishing strong theoretically based criterion variables at this stage in POM research. However as the field matures, the importance of ultimately establishing predictive validity within a theoretical network of constructs must be emphasized.

Churchill (1979) recommends that further data be gathered after the changes are made to the items in purification. Such independent samples can further validate these changes and provide additional evidence of reliability and validity. Some researchers combine these stages by doing a single round of purification. Finally, relatively newer techniques like structural equation modeling can also be performed to confirm existing measures (Gerbing and Anderson, 1988; Jöreskog, 1993). These techniques can concurrently confirm entire measurement structures for a number of variables by testing the level of 'fit' with the data.

Based on these recommendations, the following 'ideal attributes' are proposed:

(5) Are multi-item variables used?

(6) Is content validity assessed?

(7) Is field-based pretesting of measures performed?

(8) Is reliability assessed?

(9) Is construct validity assessed?

(10) Is pilot data used for purifying measures or are existing validated measures adapted?

(11) Are confirmatory methods used?

4.2. Sampling error

One of the most critical elements of the sampling procedure is the sample frame that is used to represent the population of interest. This frame may be inaccurate if it excludes elements that should be a part of the population or includes elements that should not. For instance, if a POM researcher is examining the profile of MRP systems in North American organizations and draws from a convenience sample of a group of production managers attending a conference, *sample frame error*, is introduced based on the representativeness of the frame to the population of interest (i.e., North American organizations). At the minimum, any POM survey research should describe and justify the sample frame (Fowler, 1984). Estimation of possible frame error bias (or lack thereof) by a comparison estimation of the probability of the target population being included in or excluded from the sample frame is desirable.

The next type of error involves *error of selection* which involves the error introduced if the sample used for analysis is not representative of the sample frame. Ideally, random selection from the sample

[3] Nunnally recommends Cronbach's Alphas > 0.6 for exploratory research.

frame will reduce selection error. Further, response rate should be reported to indicate the extent of the sample frame polled. If the entire sample frame responds to an instrument, there is no selection error and we have a census. However, that is usually impossible in a research context. Therefore, *high response rates* and an estimation of non-response bias should be conducted. While the higher the response rate the better, response rates of under 20% are extremely undesirable (Yu and Cooper, 1983). All efforts should be made (i.e., incentives for responding, multiple mailings, etc.) to maximize response rate. *Non-response bias* can then be estimated by sampling a group of non-respondents (or obtaining data on them through secondary sources) and comparing them with respondents on key characteristics (Armstrong and Overton, 1977). In sum, the key questions that need to be addressed are as follows:

(12) Is the sample frame defined and justified?

(13) Is random sampling used from the sample frame?

(14) Is the response rate over 20%?

(15) Is non-response bias estimated?

4.3. Internal validity error

Internal validity error addresses the question of whether differences in the dependent variable are indeed caused by the independent variable or could other variables be confounding the relationship. In experimental designs using survey research, it is possible to control extraneous effects on the dependent variable by using experimental controls or by homogenizing the sample groups. Failure or inability to do that could lead to confounding effects and erroneous conclusions. For instance, suppose a POM researcher hypothesized that use of a certain type of MRP system would be related to cost reduction. Suppose the analysis confirms this. The conclusion can lead to the recommendation of adopting this MRP system in order to achieve cost reduction. However, in the sample profiled there might have been a high correlation between adoption of that MRP system and investment in sophisticated data bases that enhance the accuracy and reliability of inventory records—the latter being the actual cause of cost reduction. The true implications are dramatically different. Therefore, POM researchers, in the absence of experimental designs, should try to justify internal validity. This can be done informally through a discussion of why causality exists or why alternate explanations are unlikely. Alternatively, more formal methods such as follow-up interviews with respondents or observation of multicollinearity among the variables can eliminate rival explanations for the findings. Structural equation modeling and other path analytic approaches (e.g., LISREL) can also be used to test causality and since these methods often test concurrent relationships among multiple variables, confidence in internal validity of the model is enhanced. Therefore, the question can be simply stated as:

(16) Are attempts made to establish internal validity of the findings?

4.4. Statistical conclusion error

Statistical conclusion error relates to the statistical power of tests being used (Straub, 1989). A greater power implies that there is a greater probability of finding statistical relationships among variables. Low power leads to erroneous conclusions. While type I error (i.e., the probability of finding a relationship when none exists) is indicated by the α level in statistical tests, the lack of power leads to type 2 errors (i.e., the probability of incorrectly sustaining the null hypothesis) (Baroudi and Orlikowski, 1989; Verma and Goodale, 1995). The power of a test is directly proportional to sample size and effect size, and inversely related to *p*-value (Sharma, 1996). However, the single most important factor in establishing adequate power for a test is sample size. Sample sizes of at least 100 are desirable, although a general heuristic for multivariate analysis is at least five times the number of variables in the model. For instance, a factor analysis of 50 items should have a sample size of at least 250 (Hair et al., 1992). In other words,

(17) Is there sufficient statistical power to reduce statistical conclusion error?

4.5. Summary discussion of sources of error

To conclude this section, it should be reiterated that any construct association in the theoretical domain needs to be adequately determined in the operational domain. There are various sources of error in this determination. These errors consist of measure-

ment error, sampling error, internal validity error, and statistical conclusion error. These errors are cumulative. For instance, a study may reduce statistical conclusion error but still have measurement, sampling, and internal validity error. Reducing sampling, internal validity, and statistical conclusion error could still lead to meaningless results if the hypothesized variables do not represent the constructs they are supposed to measure. Any survey research in POM should be cognizant of these errors and attempt to alleviate them by using prescribed attributes of conduct for this kind of research. Failure to do so can and will lead to erroneous conclusions and regression rather than progress in contribution to theory. On the other hand, careful adherence to standards can improve replicability of research, enhance the cumulative tradition of theory development, increase respectability of the field, and enhance confidence in its prescriptions to the profession. In Section 5, we use the 17 attributes listed above to evaluate the 'quality' of survey research in POM.

5. Evaluation of survey research in POM

The proliferation of field-based research in POM has come about in the past few years. An empirical study based on a survey of academicians had reported a ranking of the leading journals involved in POM research (Barman et al., 1991). The top five of these are *Management Science, Decision Sciences, Journal of Operations Management, IIE Transactions* and the *International Journal of Production Research.* Of these, the most progressive in publishing 'nontraditional' field-based work have been *Journal of Operations Management, Decision Sciences,* and *Management Science.* In addition, *Production and Operations Management* is a well respected new journal in the field that also publishes high quality POM articles. This study consequently identified and evaluated all survey research studies from these four journals conducted in the period 1990–1995 [4] as the sample of interest. We believe that this sample is representative of the leading edge quality in current field-based POM research. Another recent study that evaluated statistical power in operations management research also sampled articles from two of these four journals—*Decision Sciences* and *Journal of Operations Management* (Verma and Goodale, 1995).

A total of 25 unique articles using survey-based methodology were published in these four journals between 1990 and 1995. These articles used primary data sets that were created through surveys as a part of executing that study. Even though other empirical POM studies have been reported in this time period, they used survey data from secondary sources such as PIMS data base or used accounting and financial data from COMPUSTAT tapes. Alternatively, some of them were based on structured interviews from a set of few specific plants in which the primary researcher also spent a considerable amount of time observing actual plant operations. Case studies with a limited sample size have also been reported. Such studies were excluded since their primary data collection process did not involve the survey methodology. Multiple studies from the same data set and authors were only listed once if the scores on ideal survey attributes were identical. Studies selected for assessment are not identified by their specific citations here since the emphasis was not on rating an individual study as good or bad, but on understanding the general trends of how well on average is POM survey-based research being conducted.

An evaluation of this published POM survey-based research in these three journals along the 17 ideal survey attributes listed earlier is presented in Table 1. Each paper in our representative sample was perused in detail for the purpose of evaluating its methodological rigor. A set of key indicators was identified for each attribute in order to reduce subjectivity in the evaluation process. The presence of a '1' check mark in Table 1 indicates that the given attribute was actually considered or accounted for in that study. Conversely, a '0' check mark indicates that no consideration was given to that particular attribute. The total quality score for any published article is simply obtained by adding up the number of '1' check marks received by that study. A perfect total quality score of 17 would indicate that it possesses all the required attributes for an ideal survey-based study. No attempt is made here to give differ-

[4] *Production and Operations Management* journal started publication in 1992.

Table 1
POM survey research evaluation

Ideal survey attributes	POM empirical studies											
Type of survey research (Ref#): exploratory (XPY) or explanatory (XNY) cross sectional (CS) or longitudinal (L)	#1 XPY L	#2 XPY CS	#3 XPY CS	#4 XNY CS	#5 XNY CS	#6 XNY CS	#7 XPY CS	#8 XNY CS	#9 XPY CS	#10 XPY CS	#11 XNY CS	#12 XNY CS
General												
1. Is the unit of analysis clearly defined for the study?	1	1	1	1	1	1	1	1	1	1	1	1
2. Does the instrumentation consistently reflect that unit of analysis?	1	1	1	1	1	1	1	1	1	1	1	1
3. Is the respondent(s) chosen appropriate for the research question?	1	1	1	1	1	1	1	1	1	1	1	1
4. Is any form of triangulation used to cross validate results?	0	0	1	0	1	0	0	1	0	1	0	0
Measurement error												
5. Are multi-item variables used?	0	0	0	1	1	1	1	1	1	0	1	0
6. Is content validity assessed?	0	0	0	1	1	1	1	1	1	1	1	1
7. Is field-based pretesting of measures performed?	0	0	0	1	1	0	1	1	1	0	1	0
8. Is reliability assessed?	0	0	0	0	1	1	0	1	0	0	1	0
9. Is construct validity assessed?	0	0	0	0	1	1	0	1	0	0	1	0
10. Is pilot data used for purifying measures or are existing validated measures adapted?	0	0	0	0	1	0	0	1	0	1	1	0
11. Are confirmatory methods used?	0	0	0	0	0	0	0	0	0	0	0	0
Sampling error												
12. Is the sample frame defined and justified?	1	1	1	1	1	1	1	1	1	1	1	1
13. Is random sampling used used from the sample frame?	1	1	1	1	1	1	1	1	1	1	1	1
14. Is the response rate over 20%?	1	0	0	1	1	1	1	1	1	1	1	1
15. Is non-response bias estimated?	0	0	1	1	0	1	0	1	0	0	0	0
Internal validity error												
16. Are attempts made to establish internal validity of the findings?	0	0	0	1	1	1	0	1	0	1	1	0
Statistical conclusion error												
17. Is there sufficient statistical power to reduced statistical conclusion error?	1	1	1	0	1	1	0	0	1	1	1	1
Total quality score	7	6	8	11	15	13	9	15	10	11	14	8

ent weights to different attributes, since we believe that each attribute is extremely important in determining the eventual reliability and validity of a survey-based study. The impact of ignoring any specific attribute can significantly detract from the value of findings reported in that study. It must also be noted that several attributes typically go hand-in-hand, and thus have a cumulative effect on the final quality score. For example, measurement error attributes will generally not score well in an exploratory study where multi-item scales may not be used.

Several times the study reported did not provide enough information to evaluate whether a particular attribute was considered. It was particularly true when the survey instruments were not provided, or

#13 XNY CS	#14 XNY CS	#15 XPY CS	#16 XNY CS	#17 XNY CS	#18 XPY CS	#19 XNY CS	#20 XPY CS	#21 XNY CS	#22 XNY CS	#23 XNY CS	#24 XNY CS	#25 XNY CS	Score (%)
1	1	1	1	1	1	1	1	1	1	1	1	1	100
1	1	1	1	1	1	1	1	1	1	1	1	1	100
1	1	1	1	1	1	1	1	1	1	1	1	1	100
0	0	0	0	0	0	1	0	0	0	0	1	0	28
1	1	0	1	1	0	0	1	1	1	1	1	0	64
1	1	0	1	1	0	0	1	1	0	1	1	0	68
1	0	0	1	0	0	0	0	1	0	0	0	0	36
1	1	0	1	1	0	0	1	1	1	0	1	0	48
1	0	0	1	1	0	0	1	0	1	0	1	0	40
1	1	0	1	0	0	0	0	0	0	0	0	0	28
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	100
1	1	1	1	1	1	1	1	1	1	0	1	0	92
1	1	1	1	1	0	1	1	1	1	0	1	0	80
0	1	1	0	1	0	1	0	1	0	0	0	1	44
1	1	1	1	1	0	1	1	1	1	1	1	1	72
1	0	1	1	1	0	1	0	1	0	0	1	1	68
15	12	9	14	13	5	10	11	13	10	7	13	7	61.9

only partially shown for a few select variables. A more thorough reading was undertaken in those cases, and other related studies if cited were obtained to elicit more details about the data collection methods and analysis performed. An attribute was deemed to not have been considered if it was neither mentioned nor discussed in the main paper or any other related sources of information. Since the sample size of assessed articles was small, considerable care could be afforded in evaluating each study. In order to increase the confidence of our rating system, all the studies were evaluated by multiple individuals and ambiguous details were resolved in a consensus fashion after a detailed discussion. The rating form used for assessing published POM survey research is shown in Appendix A.

Table 2
Summary of results and prescriptions

Ideal survey attributes	Evaluation of published POM survey research[a]	Prescriptions for POM researchers
Type of survey research (Ref#): exploratory (XPY) or explanatory (XNY) cross sectional (CS) or longitudinal (L)	64% of articles are explanatory. Trend toward explanatory research. Almost all studies use cross sectional data.	· Exploratory research is important and should continue in order to understand variables in early stages of a research topic. · An increase in theoretically-based hypothesis driven studies is a trend that bodes well for the maturation of the field. · Studies and instrumentation need to build on each other more in order to contribute toward a cumulative tradition of building theory in an area. · Longitudinal studies are difficult to do, but can be very powerful in establishing causality and studying processes and should be attempted.
General		
1. Is the unit of analysis clearly defined for the study?	Very Strong (100%)	· Triangulation should be attempted in order to enhance the quality of the results.
2. Does the instrumentation consistently reflect that unit of analysis?	Very Strong (100%)	· Triangulation should include the use of multiple respondents in an organization, particularly for complex variables.
3. Is the respondent(s) chosen appropriate for the research question?	Very Strong (100%)	· Multiple methods to assess the same construct increase the confidence in measurement and should be attempted.
4. Is any form of triangulation used to cross validate results?	Weak (28%)	

Measurement error		
5. Are multi-item variables used?	Strong (64%)	• Multi-item variables must be used in both exploratory and explanatory research in order to reduce measurement error and obtain the true assessment of a construct.
6. Is content validity assessed?	Strong (68%)	• Field-based pretesting must be performed in order to clean instrumentation and establish 'real world' importance of study and constructs.
7. Is field-based pretesting of measures performed?	Weak (36%)	• All measures must be consistent within items and across respondents. Cronbach's Alpha analysis for multi-item variables at the minimum, and test–retest analysis where possible for all measures, enhances the reliability of the measure.
8. Is reliability assessed?	Moderate (48%)	• Validity of constructs, particularly convergent and discriminant validity, should be assessed in order to reduce measurement error and to improve interpretability of results.
9. Is construct validity assessed?	Weak (40%)	• To facilitate cumulative building of theory in POM, purification of measures prior to analysis or adaptation of existing valid and reliable measures is desirable and should be attempted.
10. Is pilot data used for purifying measures or are existing validated measures adapted?	Weak (28%)	• As the field matures and theoretically based measures exist, confirmatory methods to check preexisting measurement models should be conducted.
11. Are confirmatory methods used?	Very Weak (0%)	

Table 2 (continued)

Ideal survey attributes	Evaluation of published POM survey research[a]	Prescriptions for POM researchers
Sampling error		
12. Is the sample frame defined and justified?	Very Strong (100%)	• While response rates are usually over 20%, every attempt should be made to increase response from the target frame through follow-up activities.
13. Is random sampling used from the sample frame?	Very Strong (92%)	• Non-response bias should be estimated by examining data from non respondents, particularly in cases of low response rates.
14. Is the response rate over 20%?	Strong (80%)	• Generalizability or external validity of results can be enhanced by demonstrating that the sample represents the sample frame.
15. Is non-response bias estimated?	Moderate (44%)	
Internal validity error		
16. Are attempts made to establish internal validity of the findings?	Strong (72%)	• While the majority of papers discuss internal validity issues, attempts should be made to apply more formal methods such as the inclusion of control variables in analysis, assessment of multicollinearity, path analytic models, etc.
Statistical conclusion error		
17. Is there sufficient statistical power to reduce statistical conclusion error?	Strong (68%)	• While the majority of POM studies have sufficient sample size for tests, power analysis should be discussed in order to minimize erroneous conclusions.

[a]Assessment based on incidence of attribute on a 5 level scale (very strong, strong, moderate, weak, very weak) in 20 point increments from 0 to 100%.

5.1. Guidelines for methodological rigor in POM

Results of our analysis in Table 1 reveals that no study obtained a perfect quality rating score, even though a few of them came fairly close (a score of 15 out of possible 17). There was a wide range of research rigor observed, with several studies posing fairly simple and straightforward questions without much consideration being given to operationalizing multi-item constructs or focusing on reliability or validity testing. These studies typically happened to be exploratory in nature, and were conducted in relatively earlier periods. The recent studies were more explanatory in nature, and did attempt to pay more attention to using multi-item constructs that were validated through either field-based pretesting or more sophisticated statistical techniques. Due to the small sample size considered, no attempt was made to isolate any specific trends with respect to publications in different journals. However, it must be noted here that articles published in each of the four journals considered spanned a continuum of rigor.

It is interesting to observe the percentage of articles that did consider a particular attribute. This percentage score is reported in the last column in Table 1. The general attributes represented by questions 1, 2, and 3 were observed across the board, as was the definition and justification of the sample frame. Cross-validation of results through triangulation methods was observed in less than a third of the studies. Attributes related to assessment of reliability, construct validity, and purification of measures were observed in less than 50% of the studies. Finally, no studies used confirmatory methods for measurement models, partially reflecting the lack of established constructs in the field.

Overall, the pool of 25 studies assessed here considered 62% of the attributes needed for an ideal survey study. The usage was, however, not evenly distributed. POM studies suffer by-and-large from not using multiple respondents or multiple methods. Inter-rater reliability is consequently seldom assessed. This may be due to the cost and time involved in obtaining multiple responses from the same organization. In addition, robust constructs were not operationalized and several reliability related considerations were ignored.

Regarding the source of errors, we find that sampling-related errors are generally being avoided. However, non-response bias must be estimated more frequently. The majority (68%) of the studies discussed internal validity and causality issues. However, most of these issues were addressed through informal discussion of alternative explanations. Formal correlational analysis or path analytic techniques to address cause and effect were conducted in only four studies. Statistical conclusion error occurred in about 30% of the studies assessed due to the presence of small sample sizes in the data collected. Greater power of the test can be obtained by increasing the response rates. Finally, with one exception, all studies used purely cross-sectional data. While longitudinal data is difficult to collect, it serves as a powerful means of studying organizational processes and enhances confidence in the assessment of causality.

A longitudinal analysis of publication patterns reveals that 18 out of the 25 articles assessed were published in 1994 and 1995 alone. Of these 18 articles, approximately 75% were in the explanatory category. Overall for the complete data set, explanatory studies on an average possess a higher quality rating score than exploratory studies (12 vs. 8). Thus, the maturing in quality of POM survey research is being evidenced through the publication of a greater number of explanatory studies in the recent past. Such trends must be accelerated to preserve the growth of the field and bring it up to the level of rigor that is observed in other social science disciplines.

The guidelines for establishing greater rigor in POM survey research are clear from our analysis. Table 2 summarizes these prescriptions, by highlighting those aspects of survey research that require particular attention from POM researchers in order to effectively use this methodology to build a cumulative tradition of theory in the field.

6. Conclusion

The paucity of POM articles that qualified for assessment after being published in leading journals of the field is a partial indicator of how difficult it is to do high quality survey-based research in the field.

It could also indicate that the field is in a state of transition from the use of traditional modeling-based methodologies to empirical survey-based methodologies. Long lead times for data collection and analysis also lead to a time lag, whereby several years pass between the actual conceptualization of a study and its eventual publication. Thus, right now, we are reading research articles that were actually done in early 1990's, and some even towards the end of previous decade. This should be kept in mind while interpreting the implications of our findings. It is hoped that the general level of sophistication and rigor of research projects that are currently being implemented would, on an average, be higher that those that were assessed here.

We find that survey-based studies in POM should pay more attention to reporting all the details of how the data was collected, and perhaps also include the detailed questionnaire Appendix A. This would lend more credibility to the study, along with offering readers insights into how the research was conducted. Published survey-based articles in other fields of business generally present more implementation related details.

We hope that the value of this study lies in providing a usable framework for POM researchers conducting survey research. Effective research requires careful attention to both rigor and relevance. While the metric used does suffer from some subjectivity, we believe it accurately identifies the weaknesses in the conduct of POM surveys. Also, as demonstrated earlier, it is imperative that we pay close attention to these issues if we want to build coherent theory in the field. Failure to do so would lead to fragmented and isolated efforts that cannot and will not provide strong implications for practice.

As more theory building takes place through explanatory type of research models, clearly more attention will need to be paid to attributes that are related to minimizing measurement errors. In particular, we have not emphasized the use of confirmatory methods in POM even though they are widely used in other social science disciplines. In addition, greater development of psychometrically sound scales that are rooted in operations management literature is needed. A good example of this kind of research is the development by Flynn et al. (1994) of quality measurement scales that can be used by other researchers to measure quality management practices in different organizations. Thus, a cumulative tradition of building on prior research can be fostered. This would be especially true if relevant scales and constructs already available in other social science disciplines are also used in conjunction to not only compress the learning curve for the field, but also focus future research efforts on building and testing sustainable theories of production and operations management.

Appendix A. Rating form for assessing POM survey research

Ideal survey attributes	What we looked for
Type of survey research (Ref#): exploratory (XPY) or explanatory (XNY) cross sectional(CS) or longitudinal (L)	• A paper was deemed as an explanatory one if it formally stated hypothesis on the basis of prior literature or other theory/reasoning-based approaches. Studies which only focused on examining a given issue through open ended questions, or finding what is out there in the field, were considered exploratory in nature. • Studies which collected data in one point of time only were deemed cross-sectional in nature.

General	
1. Is the unit of analysis clearly defined for the study?	• A formal statement defining the unit of analysis was needed for a positive assessment on this attribute. Justification of why that unit of analysis was selected was desirable,though not considered critical.
2. Does the instrumentation consistently reflect that unit of analysis?	• The items in the questionnaire would need to be at the same level of aggregation as the unit of analysis. For example, to ensure consistency, questions pertaining to overall business strategy must have strategic business unit as the unit of analysis. In contrast, manufacturing strategy related study could have the plant as the unit of analysis.
3. Is the respondent(s) chosen appropriate for the research question?	• The person most knowledgeable at the selected unit of analysis must be the preferred respondent. It would be inappropriate for instance, to survey plant employees on organizational constructs for a multi-plant organization.
4. Is any form of triangulation used to cross validate results?	• Triangulation was judged to have been considered if more than one respondent belonging to the same unit of analysis filled out the survey questionnaire.
Measurement error	
5. Are multi-item variables used?	• Multiple items or questions would have to be used as opposed to a single item/question to define a construct of interest. A positive assessment was made if both multi-item and single item variables were used in the study.
6. Is content validity assessed?	• Content validity would need to be assessed through prior literature, or opinion of experts who are familiar with the given construct.
7. Is field-based pretesting of measures performed?	• A positive assessment was made only if the study formally stated the inclusion of this step in cleaning up the survey instrument and establishing its relevance.
8. Is reliability assessed?	• Cronbach's Alpha analysis or test–retest analysis would be needed for a positive assessment.
9. Is construct validity assessed?	• Construct validity (discriminant/convergent) analysis in the form of exploratory factor analysis, item-construct correlation, etc., would be needed for a positive assessment.
10. Is pilot data used for purifying measures or are existing validated measures adapted?	• A positive assessment was made if constructs and their associated items were evaluated on the basis of pretesting before the collection of actual data. Alternatively, con-

	structs which were well defined and tested in prior studies could also be used.
11. Are confirmatory methods used?	· Confirmatory factor analysis (e.g., using LISREL) results would need to be reported to establish construct validity.
Sampling error	
12. Is the sample frame defined and justified?	· A discussion of sample frame was needed for a positive assessment.
13. Is random sampling used from the sample frame?	· Sampling procedures (random or stratified random) would need to be discussed for a positive assessment.
14. Is the response rate over 20%?	· A formal reporting of response rate over 20% was needed for a positive assessment.
15. Is non-response bias estimated?	· A formal reporting of non-response bias testing was needed for a positive assessment.
Internal validity error	
16. Are attempts made to establish internal validity of the findings?	· At the very minimum, a discussion of results with the objective of establishing cause and effect in relationships, elimination of alternative explanations, etc., was needed for a positive assessment. Statistical analysis for establishing internal validity (like structural equation modeling) was considered as desirable, but not critical.
Statistical conclusion error	
17. Is there sufficient statistical power to reduce statistical conclusion error?	· At least a sample size of 100 and an item to sample size ratio of more than 5 were needed for a positive assessment.

References

Armstrong, J.S., Overton, T.S., 1977. Estimating nonresponse bias in mail surveys. J. Marketing Res. 14, 396–402.

Babbie, E., 1989. The Practice of Social Research. Wadsworth, Belmont, CA.

Bagozzi, R.P., 1979. The role of measurement in theory construction and hypothesis testing: toward a holistic model. In: Conceptual and Theoretical Developments in Marketing. Am. Marketing Assoc. Chicago, IL.

Barman, S., Tersine, R.J., Burley, M.R., 1991. An empirical assessment of the perceived relevance and quality of POM-related journals by academicians. J. Operations Manage. 10 (2), 194–212.

Baroudi, J.J., Orlikowski, W.J., 1989. The problem of statistical power in MIS research. MIS Q. 13, 87–106.

Campbell, D.R., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull. 56, 81–105.

Churchill, G.A., 1979. A paradigm for developing better measures of marketing constructs. J. Marketing Res. 16, 64–73.

Cote, J.A., Buckley, R.M., 1987. Estimating trait, method, and error variance: generalizing across 70 construct validation studies. J. Marketing Res. 24, 315–318.

Cote, J.A., Buckley, R.M., 1988. Measurement error and theory testing in consumer research: an illustration of the importance of construct validation. J. Consumer Res. 14, 579–582.

Dubin, R., 1978. Theory Building. Free Press, New York.

Eisenhardt, K., 1989. Building theories from case study research. Acad. Manage. Rev. 14 (4), 532–550.

Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K.A., Flynn, E.J., 1990. Empirical research methods in operations management. J. Operations Manage. 9 (2), 250–284.

Flynn, B.B., Schroeder, R.G., Sakakibara, S., 1994. A framework for quality management research and an associated measurement instrument. J. Operations Manage. 11, 339–366.

Fowler, F.J., 1984. Survey research methods. Appl. Social Res. Meth. Series, Vol. 1. Sage Publications, Beverly Hills, CA.

Gerbing, D.W., Anderson, J.C., 1988. An updated paradigm for scale development incorporating unidimensionality and its assessment. J. Marketing Res. 25, 186–192.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1992. Multivariate Data Analysis with Readings. Macmillan, New York.

Huber, G.P., Power, D.J., 1985. Retrospective reports of strategic-level managers: guidelines for increasing their accuracy. Strategic Manage. J., 171–180.

Hunter, J.E., Schmidt, F.L., Jackson, G.B., 1983. Meta-analysis: cumulating research findings across studies. Sage, Beverly Hills, CA.

Jöreskog, K.G., 1993. Testing structural equation models. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. Sage Publications, Newbury Park, CA.

Kerlinger, F.N., 1986. Foundations of behavioral research. Holt, Rinehart & Winston, New York.

McGrath, J., 1979. Toward a theory of method for research on organizations. In: Mowday, R.T., Steers, R.M. (Eds.), Research in Organizations: Issues and Controversies. Goodyear Publishing, Santa Monica, CA, pp. 4–21.

Nunnally, J.C., 1978. Psychometric Theory. McGraw-Hill, New York.

Sharma, S., 1996. Applied Multivariate Techniques. Wiley, New York, pp. 349–350.

Simon, H., 1980. The behavioral and social sciences. Science 209, 72–78.

Straub, D.W., 1989. Validating instruments in MIS research. MIS Q. 13, 147–169.

Verma, R., Goodale, J.C., 1995. Statistical power in operations management research. J. Operations Manage. 13, 139–152.

Yu, J., Cooper, H., 1983. A qualitative review of research design effects on response rates to questionnaires. J. Marketing Res. 36, 36–44.