

# 1

## INTRODUCTION

### THE PURPOSE OF THE BOOK

Empirical science in operations management (OM) has advanced dramatically since the early 1990s. This advancement has been driven largely by scholarly inquiry into manufacturing and service operations strategy, and subsequently in areas such as quality management, just-in-time (JIT), supply chain management, service operations, and e-operations. These topics are inherently less understood, boundary spanning, and relatively unstructured in contrast to more traditional inventory theory, for example. As a consequence, the time was ripe to open the OM field to new approaches and methods of scholarly inquiry. The era of OM empiricism began. Yet studying complex OM phenomena, typically through empirically based survey and case research methods, oftentimes requires the development of new constructs and perceptual measures that define them. Moreover, the use of objective metrics that are drawn from secondary data sources necessitates a reasonable assessment of data quality.

A primary tenet of good empirical science is this: *Reliable and valid metrics are vital to the advancement of theory.* The pioneers of OM empirical research have had little guidance from their own discipline for developing appropriate metrics. Few early OM scholars had expertise in

the social science and the statistical foundations of multi-item measurement scale development. Thus, to improve the theory and practice of OM, a deeper understanding of items and scale construction of perceptual measures was needed. No one study makes a measurement scale “valid.” Validity first requires that metrics be reliable in that they are consistent, and second, it requires replication and refinement of existing metrics to adequately glean the meaning of the intended constructs and their posited interrelationships.

This handbook is a compilation of multi-item measurement scales and objective metrics that have been used to construct surveys and questionnaires in published, empirical OM research and/or have been derived from secondary data sources. Some of these constructs have also been deployed in case research investigations, and more recently in behavioral operations experiments. We envisioned this handbook as providing a broad-based picture of the empirical side of the field and serving as a springboard for stimulating future OM empirical research.

Historically, research in OM employs both analytical models and empirical science. The empirical science applications in OM are relatively new compared with analytical modeling, which traditionally dominated the field’s research landscape. The past two decades witnessed a significant rise

## 2 • HANDBOOK OF METRICS FOR RESEARCH IN OPERATIONS MANAGEMENT

in prominence of the empirical side of OM scholarship, which builds on social, economic and behavioral science approaches. Much of the empirical work to date employs perceptual metrics that consist of multiple-item scales as well as objective measures. Yet the OM field lags the more mature management disciplines, such as marketing, management, strategy, and information systems, which employ rigorous social and behavioral science methods in the conduct of empirical research. It also falls short of economics, finance, and accounting in the use of objective measures that are frequently derived from publicly available secondary sources, such as COMPUSTAT, CRSP, Hoovers, and EdgarScan.

Until recently, there were few metrics available that operationalize OM concepts and constructs; and there was little experience from which to build the substantive areas of our discipline. Many of the original measures used in pioneering OM research were developed from scratch or were drawn from research in other disciplines. As the OM field matures and theories and frameworks are empirically tested, higher quality perceptual and objective metrics are required if the resulting research is to be useful for explanation, prediction, decision making, and testing of assumptions.

We conceptualized this handbook as a source in which the body of empirical studies in OM could be mapped via the constructs and multi-item measures that tap into them. As the amount of empirical research in OM expands, it has become increasingly obvious that many of the metrics employed are being redeveloped over and over again, and oftentimes, different constructs use subsets of the same items. This lack of semantic clarity in our constructs and reuse of the same items in scales that are intended to capture different constructs pose serious threats to validity. Invalid measures can have a dire impact on the ability of OM to develop sound theory in the longer term.

Our reviews indicate that the accumulated experience of iterative development has not been systematically summarized and mistakes in measurement scale development are repeated in exploratory modes of investigation. Perceptual metrics in OM research are not being refined and honed for different contexts as required for good empirical research and their quality varies widely. For example, Rosenzweig and Roth (2004) report that no

adequate multi-item scales were available for testing the elements of competitive progression theory. Since this handbook summarizes the existing set of perceptual metrics in the OM field, others can learn from the pioneers and, in turn, can begin the iterative multi-item, scale development that over time will have the high degrees of reliability and validity needed for theory building and testing. Similarly, many objective items, which are used as proxy variables for theoretical constructs and performance outcomes, are offered for others to build on.

The multi-item scales presented here can be used to measure constructs that have a large OM empirical base, including total quality management (TQM), JIT or lean manufacturing, operations capabilities and strategy, technology, operations supply chain management, performance of operations, as well as emerging areas such as new product development, service operations, environmental and sustainable operations, and cross-functional and organizational interfaces. Thus, this handbook is a broad compilation of the literature and measures that serve as reference work to provide researchers a guide for construction of surveys and other structured data collection on OM topics. It sets the stage for the second generation of OM empirical science, that is, it builds on more confirmatory, as opposed to first generational exploratory approaches to measurement.

Our book is designed to be a basic reference source for furthering scholarly research in OM. Among others, this reference book is particularly useful to four key groups: (1) OM professionals conducting survey and other empirical research or teaching empirical science to graduate students; (2) researchers wishing to conduct meta-analyses of the psychometric properties of items and measurement scales; (3) OM faculty and doctoral students who are interested in empirical research and desire to become familiar with available constructs and measurement instruments; and (4) scholars in other fields who are interested in investigating OM-related research, especially regarding the interfaces of OM with other disciplines and across business and supply chain entities. Researchers now have ready access to the body of existing multi-item scales and objective items published in the extant OM literature and their known psychometric properties as reported by the authors.

## COVERAGE AND SCOPE

---

Since our focus is to identify and compile a body of knowledge on relevant multi-item measurement scales and objective measures, we began our search with articles starting in 1990, which is generally recognized as the year when empirical research in OM was taking off. To find them, we first reviewed every issue of production and operations management related journals: *Journal of Operations Management (JOM)*, *Production and Operations Management (POM)*, *Manufacturing and Service Operations Management (M&SOM)*, *Decision Sciences (DS)*, *Journal of Service Research (JSR)*, *International Journal of Production Research (IJPR)*, *Management Science (MS)*, and *International Journal of Operations and Production Management (IJOPM)*. The first four journals (*JOM*, *POM*, *M&SOM*, and *DS*) were searched manually. The last four journals (*JSR*, *IJPR*, *MS*, and *IJOPM*) were reviewed through computer databases (i.e., ABI and EBSCO). In addition, we used a keyword search of the following terms: *empirical and operations*, *empirical and manufacturing*, and *empirical and service*. We found at least one article in other journals, including *Academy of Management Journal (AMJ)*, *Strategic Management Journal (SMJ)*, *Technovation*, and *Resource Policy*. We also supplemented these with multi-item measurement scales found in selected book chapters. While we have attempted to be as thorough as possible in our coverage of important OM scales, we do not claim that this book covers every multi-item measurement scale in OM to date. Articles that do not give sufficient information on measurement, such as the lack of a scale listing, were excluded. Our search produced 230 journal articles in the literature that have published one or more OM-related scales. Additionally, we illustrate questions that employ single-item objective measures in Chapter 3. In total, 1,803 multi-item measurement scales and numerous objective measures have been found.

To the extent that the authors provided the information, each article has been summarized in a standard format consisting of the following elements:

*Citation:* The full citation for the article being summarized is provided.

*Scale listing:* A list of the questions (items) included in each of the scales or objective measures tested in the article is given.

*Construct description:* A description of the meaning of the scales, the literature used, and grounding of the constructs are provided. A further description of the items is also included if presented in the original journal article.

*Measurement description:* We summarize (1) how measurement was performed, (2) whether the scales were grouped into factors, and (3) what format was used (e.g., Likert) for the scaling of the questions.

*Development:* A description of the measurement and survey development process is summarized. This covers the process used to develop the scales, including pretesting, split-sample approaches, exploratory factor analysis, or other methods.

*Sample:* The nature of the sample employed is given. This part includes the following information as indicated by the author(s) of the manuscript: the data collected to demonstrate reliability and validity; the survey respondents (e.g., managers, nurses, students); the sample size and the response rate; the measures taken to reach nonrespondents and to test for possible response bias.

*Reliability and validity:* The psychometric properties of the tests are summarized as reported in the original manuscript. The summary covers (1) the measures reported to test for reliability of the scale (e.g., Cronbach alpha, split halves methods, or measurement model statistics derived from confirmatory factor analyses); (2) tests used for convergent and discriminant validity; and (3) the demonstration of unidimensionality.

*Scores:* The methods used to obtain a scale score are provided if given in the text. Most frequently used methods are the averaging of items comprising the scale or the use of factor loadings. Any other scoring information that the authors of the manuscript made available is also reported.

*Other evidence:* Any other evidence related to the measurement reliability and validity of the scale is given.

*References:* Citations for the references used in the summary are reported.

## CAVEATS AND CAUTIONS

---

The criteria above are given to provide the reader with some notion about the constructs, items, and scales as reported in the original source

#### 4 • HANDBOOK OF METRICS FOR RESEARCH IN OPERATIONS MANAGEMENT

materials. They are intended to offer some insights into the psychometric properties of the scales, but the measurement quality cannot be guaranteed since we provide only a simple description of the psychometric properties as reported by the authors. The measurement quality of reported OM metrics varies widely. Many potentially interesting and insightful OM empirical articles have paid inadequate attention to and demonstrated a lack of rigor in the development of perceptual metrics used in the study design. Furthermore, efforts to rectify inadequate metrics after the full study has been implemented have been problematic in that many scales have been modified on a post hoc basis without further replication on a separate validation or holdout sample. By identifying these items and scales, OM researchers can replicate and refine them in the same and different contexts.

We also caution that it is difficult to interpret and compare the quality of the metrics because of differences in the statistical procedures, contexts, and response rates reported in the original source material. Oftentimes, the methods information is insufficient for study replication. Moreover, some authors may not have fully disclosed all the limitations of their research. We, therefore, urge researchers interested in particular scales to review the original sources for obtaining a deeper understanding of the scales and the context of the original research and to assess the theoretical rationale for constructs and items comprising their operationalization.

Also, for most of the literature-based measurement scales in OM, and especially those in the early years of OM empirical research or for emerging topics, the psychometric properties may not be consistently reported. As indicated earlier, our review discovered a lack of semantic clarity and well-defined, construct definitions. Similar items tapping into one construct in some studies will be reported in different constructs in another, which pose a threat to validity. Future research will be required to rectify these conceptual and operational issues. Consequently, the metrics here should be viewed as a *starting point* for further refinement and scale assessment.

We recognize that the literature in multi-item, measurement scale development is rich and varied. Much of the early work in scale development

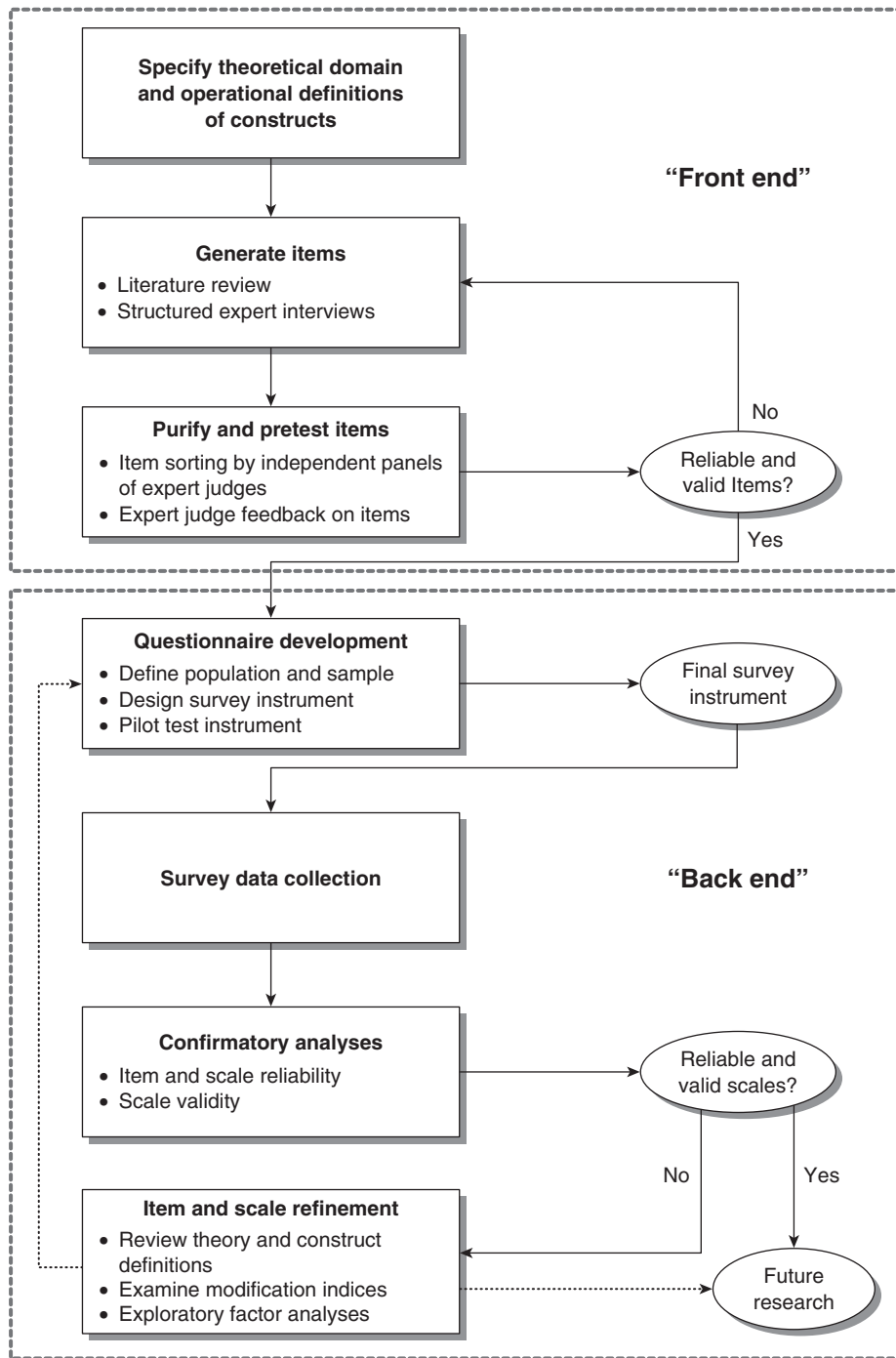
was based on exploratory, first-generation multivariate techniques, such as principal components analysis or common factor analyses. Some studies used a combination of exploratory procedures and confirmatory data analysis on the same samples, which is not recommended because of data fitting. There are many useful statistical approaches for developing new perceptual metrics in the extant literature that can add to the second generation of OM empirical science. Building on Churchill (1979), Menor and Roth (2007) offer a two-stage approach for developing and assessing items and measurement scales (see Figure 1.1). In contrast to many perceptual scale development efforts, these authors suggest that researchers focus on the “front-end” stage of survey research—*before* pilot testing and field implementation of questionnaires—for tentative item and scale assessment. They demonstrate that attention to the front end can yield an efficient and effective means of improving overall measurement quality. Figure 1.2 illustrates the Menor and Roth (2007a, 2007b) statistics that can be used for establishing tentative reliability and validity at the front-end stage. Examples of the Menor and Roth approach can be found in Stratman and Roth (2002), Menor and Roth (2007b), and Rosenzweig and Roth (forthcoming).

Froehle and Roth (2004) provide yet another rigorous approach for addressing the front end of multi-item scale development when samples are large, and Froehle and Roth (2007) offer a new methodology for developing and defining constructs. In addition, comparison of the pros and cons of various survey data collection methodologies for OM researches can be found in Rosenzweig, Roth, and Gilland (2006).

A selected bibliography listed at the end of this chapter reports on issues that need to be considered in a careful construction of new items and a rigorous analysis of perceptual data that will increase the rigor of the researcher’s efforts to further develop and refine psychometrically sound multi-item scales.

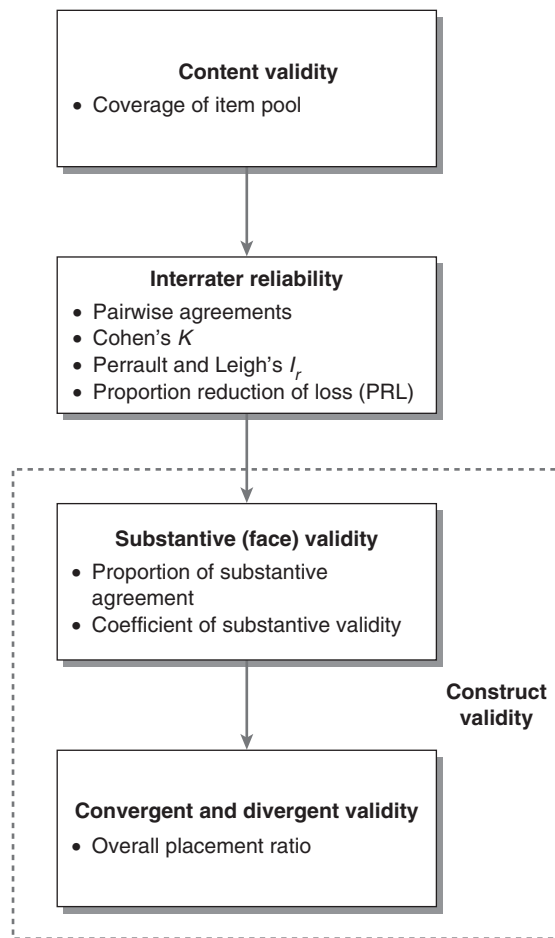
#### ORGANIZATION OF THE BOOK

Chapter 2 offers a classification scheme for the scales. This method is unique to this book and provides a typology for scales in OM. It is based



**Figure 1.1** Menor and Roth's Two-Stage Approach for New Item and Scale Development

## 6 • HANDBOOK OF METRICS FOR RESEARCH IN OPERATIONS MANAGEMENT



**Figure 1.2** Statistics for Establishing Tentative Reliability and Construct Validity at the Front End

SOURCE: Menor and Roth (2007a, 2007b).

on the classification systems used for research topics accepted by major journals in the OM field.

Chapter 3 discusses objective, single-item metrics that have been covered in OM research. It also includes illustrative questions from five sample surveys used in OM. The unit of analysis is theoretically and pragmatically important to OM empirical research, and it varies across studies. In Chapter 3, three of the surveys highlight questions at the “plant” level, and two represent the “operations business unit” level of analyses, one from manufacturing and the other from service. Less frequent, but

also of increasing importance, to understanding the customer/buyer side of operations, is the use of “individuals” as the unit of analysis (*see, e.g.,* Froehle and Roth 2004; Siemsen, Balasubramanian, and Roth, forthcoming).

Chapter 4 contains 230 summaries of journal articles arranged in alphabetical order by last name of the first author. These summaries are useful for those looking for a particular scale in the published literature.

A particular scale of interest can be found in three ways by using the appendixes to this book.

If only the general category of the scale is known (e.g., manufacturing strategy, JIT, quality management), one or more scales can be found in Appendix 1. On the other hand, if the specific scale name is known (e.g., employee involvement), it can be found in Appendix 2. If the author of the scales is known, the location of the relevant article summary can be found in the author listing in Appendix 3 and in the alphabetical table of contents for Chapter 4.

#### SELECTED BIBLIOGRAPHY OF METHODS FOR SCALE DEVELOPMENT

- Anderson, J. C., and D. W. Gerbing. 1991. Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology* 76 (5): 732–40.
- Blalock, H. M. 1979. Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Journal of Sociological Review* 44:881–94.
- Bollen, K. A. 1989. *Structural equations with latent variables*. New York: Wiley.
- Campbell, D. T. 1960. Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist* 15:546–53.
- Campbell, D. T., and D. W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81–105.
- Churchill, G. A. 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 6:64–73.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–46.
- Cortina, J. M. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 78 (1): 98–104.
- Crittenden, K. S., and R. J Hill. 1971. Coding reliability and validity of interview data. *American Sociological Review* 36:1073–80.
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334.
- Drolet, A. L., and D. G. Morrison. 2001. Do we really need multiple-item measures in service research? *Journal of Service Research* 3 (3): 196–204.
- Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76 (5): 378–82.
- Forza, C. 2002. Survey research in operations management: A process-based perspective. *International Journal of Operations & Production Management* 22 (2): 152–94.
- Froehle, C., and A. V. Roth. 2004. New measurement scales for evaluating perceptions of the technology-mediated customer experience. *Journal of Operations Management* 22 (1): 1–21.
- Froehle, C., and A. V. Roth. 2007. A resource-process framework of new service development. *Production and Operations Management*, 16 (2): 169–188.
- Frolich, M. T. 2002. Techniques for improving response rates in OM survey research. *Journal of Operations Management* 20:53–62.
- Gerbing, D. W., and J. G. Hamilton. 1996. Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling* 3:62–72.
- Handfield, R. B., and S. A. Melnyk. 1998. The scientific theory-building process: A primer using the case of TQM. *Journal of Operations Management* 16: 321–39.
- Harmon, H. H. 1976. *Modern factor analysis*. 3rd ed. Chicago, IL: University of Chicago Press.
- Hensley, R. L. 1999. A review of operations management studies using scale development techniques. *Journal of Operations Management* 17:343–58.
- Menor, L., & A. V. Roth, 2007a. *Improving perceptual measurement in operations management survey research*. Toronto, Ontario, Canada: Ivey Business School, University of Western Ontario.
- Menor, L. and A. V. Roth. 2007b. New service development competence in retail banking: Construct development and measurement validation. *Journal of Operations Management*, 25 (4): 825–846.
- Meredith, J. R., A. Raturi, K. Amoako-Gyampah, and B. Kaplan. 1989. Alternative research paradigms in operations. *Journal of Operations Management* 8 (4): 297–326.
- Miller, M. B. 1995. Coefficient alpha: A basic introduction from the perspective of classical test theory and structural equation modeling. *Structural Equation Modeling* 2 (3): 255–73.
- Mitchell, S. K. 1979. Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin* 86 (2): 376–90.
- Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric theory*. New York: McGraw-Hill.
- O’Leary-Kelly, S. W., and R. J. Vokurka. 1998. The empirical assessment of construct validity. *Journal of Operations Management* 16:387–405.

## 8 • HANDBOOK OF METRICS FOR RESEARCH IN OPERATIONS MANAGEMENT

- Perreault, W. D., and L. E. Leigh. 1989. Reliability of nominal data based on qualitative judgments. *Journal of Marketing Research* May:135–48.
- Reise, S. P., K. F. Widaman, and R. G. Pugh. 1993. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin* 114 (3): 552–66.
- Rosenzweig, E., and A. V. Roth. 2004. Towards a theory of competitive progression: Evidence in high tech manufacturing. *Production and Operations Management* 13 (4): 354–68.
- Rosenzweig, E., and A. V. Roth. Forthcoming. B2B seller competence: Construct development and measurement using an operations strategy lens. *Journal of Operations Management*.
- Rosenzweig, E. D., A. V. Roth, and W. Gilland. 2006. *Web-based survey best practices for the operations management researcher toolkit*. Atlanta, GA: Goizueta Business School, Emory University.
- Roth, A. V. Forthcoming 2007. Special issue on applications of empirical science in operations management. *Manufacturing and Service Operations Management*.
- Rungtusanatham, M. 1998. Let's not overlook content validity. *Decision Line* July:10–13.
- Rust, R. T., and B. Cooil. 1994. Reliability measures for qualitative data: Theory and implications. *Journal of Marketing Research* February:1–14.
- Schmitt, N., and D. Stults. 1986. Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement* 10:1–22.
- Schriesheim, C. A., K. J. Powers, T. A. Scandura, C. C. Gardiner, and M. J. Lankau. 1993. Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management* 19 (2): 385–417.
- Schwab, D. P. 1980. Construct validity in organizational behavior. *Research in Organizational Behavior* 2:3–43.
- Siemsen, E., S. Balasubramanian, and A. V. Roth. Forthcoming. Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science*.
- Singleton, R. A., B. C. Straits, and M. M. Straits. 1993. *Approaches to social research*. 2nd ed. New York: Oxford University Press.
- Stephenson, W. 1953. *The study of behavior: Q-Technique and its methodology*. Chicago, IL: University of Chicago Press.
- Stratman, J. K., and A. V. Roth. 2002. Enterprise resource planning (ERP) competence constructs: Two-stage multi-item scale development and validation. *Decision Sciences* 33 (4): 601–28.
- Swamidass, P. M. 1991. Empirical science: New frontier in operations management research. *Academy of Management Review* 16 (4): 793–814.