SO YOU ALREADY HAVE A SURVEY DATABASE?—A SEVEN-STEP METHODOLOGY FOR THEORY BUILDING FROM SURVEY DATABASES: AN ILLUSTRATION FROM INCREMENTAL INNOVATION GENERATION IN BUYER–SELLER RELATIONSHIPS

SUBROTO ROY
University of New Haven

Across business disciplines, the importance of database research for theory testing continues to increase. The availability of data also has increased, though methods to analyze and interpret these data lag. This research proposes a method for extracting strong measures from survey databases by a progression from qualitative to quantitative techniques. To test the proposed method, this study uses the Industrial Marketing and Purchasing (IMP) survey database, which includes data from firms in several European countries. The proposed method consists of two phases and seven steps, as illustrated in the context of the firm's incremental innovation generation for buyer–seller relationships. This systematic progression moves from a broad but valid empirical case study to the development of a narrow and reliable measure of incremental innovation generation in the IMP database. The proposed method can use supply chain survey databases for theory development without requiring primary data collection, assuming certain conditions.

Keywords: *secondary data; survey databases; supply chain; innovation; case study*

INTRODUCTION

The use of database research to test theory has grown increasingly common in business research, largely because of the advantages offered by business databases, such as convenient access to data, easier replication of the research and the possibility of global research teams working with the same database across time zones and in different countries. Business databases also allow for the use of multivariate statistics in theory testing, which apply rigorous quantitative methodology to business research without the enormous logistical burden of data collection. Business databases may include surveys (e.g., Institute for Supply Management^{TM}'s [ISM's] Report on Business, Industrial Marketing and Purchasing [IMP] databases) or secondary data generated from financial (e.g., COMPUSTAT) or other legal (e.g., U.S. patent) reports. Existing survey databases are the subject of this article.

The most common type of surveys used in supply chain research are "one-off" surveys that investigate a particular research question of interest. These one-off surveys, although using credible lists like the ISM member list, suffer from common problems. A typical methodology section from such research, therefore, would read:

> Any survey in supply chain management faces a *difficult task of routing the instrument* to the appropriate person in an organization, since a supply chain encompasses many units within an organization. Although the cover letter specifically requested to forward the questionnaire to other person(s) within the organization if the recipient was not the appropriate person, it was reported (via personal contact) that *many questionnaires were lost/misplaced* during the transmittal process. As a result, 177 returns were received out of approximately 1,800 questionnaires (9.8 percent). Of the 177 returns, six returns were deleted due to incomplete information, leaving 171 usable returns for analysis (9.5 percent). *Such a seemingly low rate of return is not unusual*. For example, Rutner and Gibson (2001) reported in their study on a logistics information systems survey that different survey

The author acknowledges the inputs of his Ph.D. Advisor, Professor Ian F. Wilkinson, now at the University of New South Wales, to the earlier Ph.D. dissertation version of the article.

> techniques yielded different rates of return ranging from 3.7 percent with a Mail-Fax survey to 12.6 percent with pre-called mail surveys with an average yield rate of 6 percent (Kwon and Suh 2004, p. 8, emphasis added).

This problem can be mitigated substantially by institutionalized survey databases (e.g., U.S. Census, ISM Report on Business) or by those whose data collection involves substantial resources and time expended by academic investigators (e.g., IMP databases).

Still, the use of survey databases entails certain challenges. By definition, collected data reflect the theoretical framework of the original database designers, who might not have been able to foresee all the data requirements of subsequent research questions. In addition, database research features self-selecting respondents, units of observation and data cleaning (Marshall and Buzzell 1990), so though databases offer quantitative rigor, researchers face the unpleasant prospect of sacrificing relevance because they did not create the database themselves for their particular research. Despite these criticisms, database research clearly continues to increase (Scandura and Williams 2000), and along with the proliferation of data in the Internet age, makes it important to develop better methods for extracting meaning from these databases.

Despite calls for greater attention to the development of effective methods in database research (Sheth 1996; Scandura and Williams 2000), no published accounts suggest how researchers might balance the need for relevance against the demands for rigor. This gap is surprising for three theoretical and one practical reason. First, the research trend that focuses only on relevance (through qualitative studies) is subject to considerable criticism (Scandura and Williams 2000). Second, the lack of validity and the unreliability of the measures in quantitative research also provokes lament (Peter 1979). Third, several researchers had called for increased nomological or theoretical validity in business research (Churchill 1979; Gerbing and Anderson 1988). Fourth and practically, given the proliferation of the Internet and electronic data collection methods, data keep increasing but without meaningful analyses because of the absence of a methodology that can derive meaning from those growing databases. This study therefore proposes a progression, from qualitative to quantitative techniques that may help extract meaning from survey databases.

The proposed method consists of a systematic progression, from the broad to the narrow and from qualitative to quantitative, rather than an attempt to reach convergence, such as through triangulation. Whereas triangulation involves multiple methods that converge to offer similar results (e.g., Colgate 1998), the proposed method allows for systematic focusing of the database measures on the particular theoretical context chosen by the researcher. The theoretical context investigated herein pertains to the innovation generation construct in the context of global supply chain relationships.

The rest of this article is structured as follows: The next section provides an overview of the proposed methodology, followed by an illustration of each step of the proposed seven-step method in the context of innovation generation, which consists of two main phases, as shown in Figure 1. Finally, the last section presents a checklist for applying the proposed methodology, along with a discussion of the contributions to theory and practice.

THE SEVEN-STEP METHODOLOGY

In Phase I (Steps 1–4) of the methodology, following Ghauri and Gronhaug (2005), the researcher considers the research domain and looks for potential matches with constructs from the database. This phase does not include any effort to analyze the data from the database; rather, the researcher articulates the theoretical lens that will guide the entire project. The final step (Step 4) of the first phase involves a qualitative investigation of a case study that speaks to the theoretical domain of interest and the questionnaire items available in the database.

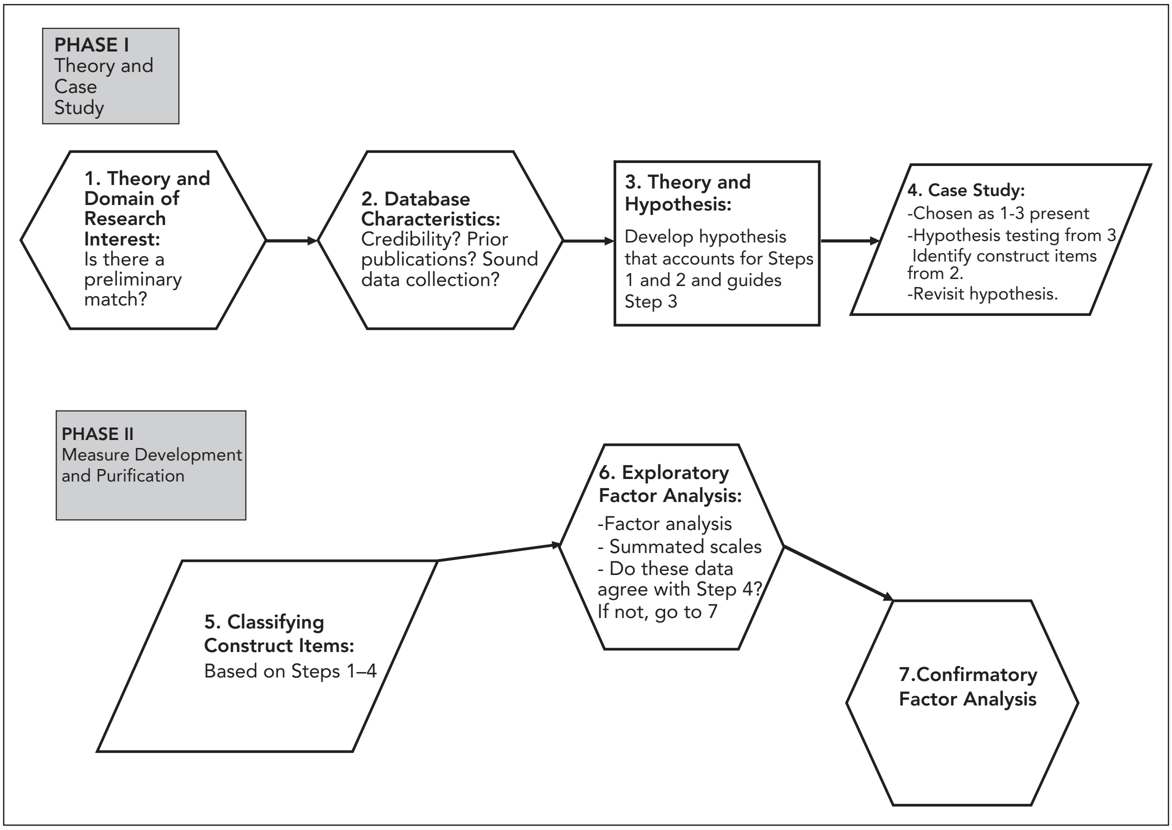
In Step 1, the researcher starts reviewing the theoretical area of research interest and the kinds of dependent (e.g., incremental innovation generation) and independent (e.g., buyer–seller relationships) variables to use. Some preliminary hypotheses about the variables of interest should emerge during this step. The researcher also completes an extensive literature review, including a review of measurement issues associated with the constructs of interest.

In Step 2, the researcher examines the structure of and items in the database and decides whether, *prima facie*, the construct items of interest appear available. In the illustration herein, the IMP database contains items that appear to measure innovation generation in buyer–seller relationships. Assuming that at least some relevant construct items appear available in the database, the researcher proceeds to Step 3. Moreover, it may be possible to pool data from various sources and for different measures over time to conduct longitudinal analysis. Therefore, in addition to finding construct measures of behavior in survey databases, the researcher might obtain firm financial data from COMPUSTAT, e.g., Sorescu, Chandy and Prabhu (2007) offer a good example of the use of multiple archival databases.

In Step 3, the researcher articulates potential hypotheses and a literature-based justification. The theoretical lens thus becomes clearer before the qualitative case study begins (Yin 2008). The more focused the theoretical review and literature-based hypothesis development is, the better the outcome of the next step will be (Eisenhardt 1989).

Finally, Step 4 at the end of Phase I involves the qualitative case study. The researcher identifies a case study site at which the phenomena of research interest have

FIGURE 1
A Seven-Step Methodology of Theory Building from Survey Databases, with an Illustration of Incremental Innovation Generation in Buyer–Seller Relationships



Note: Bold lines indicate relationships illustrated herein.

occurred and follows established case study methods (Yin 2008), though with one modification. That is, database items must be understood in the context of the case situation, and managers close to the phenomena should provide feedback about which items seem most pertinent to their situation. After gaining an understanding of the structure of the phenomena of interest, the researcher administers the items of the questionnaire to knowledgeable respondents as part of the in-depth interviews. In addition, the researcher identifies empirically valid examples of various constructs and attempts to determine their dimensions, as well as how they relate in the context of the specific case study. Finally, the researcher revisits the initial theoretical model and hypotheses and modifies the relationships among the various constructs, in accordance with the findings of the case study. With this clearer focus on the hypotheses and better understanding through the case study, the researcher initiates the second phase of data analysis.

Phase II deals with the actual data, their preparation and their purification. Step 5 involves the steps from Phase I to identify questionnaire items that reflect each construct, in the context of the theory being investigated. For example, the IMP questionnaire items (item pool) should reflect incremental innovation generation. At this step, the researcher hopes to include all items of the questionnaire that may reflect the construct, as verified by the case study (Step 4) and guided by previous theoretical work. When these data appear in the statistical package, the data preparation begins. During this particularly sensitive step, the researcher must account for any differences between the descriptive data and previously published work based on the same database, such as variations in the number of firms or demographic data (e.g., average sales), as well as any changes in the description of the database parameters. The researcher also must recode the question responses to be consistent with the hypotheses, as discussed subsequently.

For Steps 6 and 7, the researcher follows the standard statistical processes that precede multivariate analysis (Hair, Anderson, Tatham and Black 1998) in one-off surveys. In Step 6, exploratory factor analysis (EFA) should refine the items, such that those items that load on one factor reflect subconstructs of the main construct. The researcher also revisits any case study notes to confirm that the factor items make sense as subconstructs. Finally, the items can be summed together, with the Cronbach's α providing evidence of the reliability of the developed scale.

If more than one dimension of the construct appears in Step 6, Step 7 features an examination of the construct structure through confirmatory factor analysis (CFA) procedures using structural equation modeling. If there are sufficient observations, the researcher may follow a split sample approach and conduct EFA on half the sample and CFA on the hold-out sample (e.g., Houghton and Neck 2002). However, as this methodology conducts EFA preceded by strong theory development and empirical case study (Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg and Williams 1997), the CFA may be carried out on the same data set without fear of capitalization on chance (following Van Prooijen and Van Der Kloot 2001; Shi, Kunnathur and Ragu-Nathan 2005).

PHASE I: THEORY AND CASE STUDY

Theory and Domain of Research Interest

Incremental innovation generation (Popadiuk and Choo 2006) refers to changes to a dominant design or technology platform (Sood and Tellis 2005), which often occurs in buyer–seller relationships (Roy, Sivakumar and Wilkinson 2004; Pisano 2006). Existing literature seldom discusses incremental innovation without comparing it with radical innovation (see Popadiuk and Choo 2006) to clarify the contributions of innovation research. This balanced approach in both conceptual and empirical work suggests that to extend knowledge, researchers should clearly position their database analyses in the context of both radical and incremental innovations. Therefore, the database characteristics also require detailed consideration.

TABLE I

Structure of IMP Questionnaire

Section	Title
I	The focal relationship
I.1	Development
	Start and duration
	Development pattern
I.2	Characteristics of the exchange
	Seller's offer
	Importance
	Need
	Transactions
	Adaptations
I.3	Organizational setting
I.4	Relationship atmosphere
II	Connected relationship
	Relationships connected to the customer company
	Relationships connected to the supplier company
	Relationships connected to the intermediary
III	Interacting parties
III.1	The supplier
	Structural data
	Market and competition
	The selected national market
	Technology
III.2	The intermediary
	Structural data
III.3	The customer
	Structural data
	Market and competition

Database Characteristics

A database should have several characteristics to support the method suggested herein. First, it should contain data points that reflect the research interest. The IMP database, e.g., exhibits preliminary promise in terms of the interest of the researcher, as outlined in Step 1. The "Adaptations" (see Table I) section in particular contains question items that appear to address incremental innovation generation.

Second, the researcher should well understand the original purpose of the data collection. In the focal database, the network of IMP scholars wanted to collect interfirm data to test their early and then-revolutionary theory (Hakansson and IMP Group 1982) that businesses are interconnected and do not operate as islands. Their aim was to establish a comprehensive database of interfirm relations that could be used to examine various issues associated with business relations and networks. Several other publications have resulted using this database, which was compiled toward the end of the 1990s (e.g., Blankenburg-Holm, Eriksson and Johanson 1996, 1999; Wiley, Wilkinson and Young 2006).

Third, the data collection method should be clear and credible. The IMP data were collected through in-depth, personal interviews conducted by IMP member professors based in the country of the firm being interviewed. Thus, German professors interviewed German companies, French professors interviewed respondents from French companies and so on. The subsequent extension of the study to various Asian countries, including China, used the same research instruments. Interviews in Europe featured marketing executives who selected one of their firm's most important customers in a specific country to ensure an even distribution of relationships across customer countries. The sample of relationships investigated emerged as evenly distributed across seven customer countries (France 23.5 percent, United Kingdom 16.2 percent, Germany 14.0 percent, Sweden 14.0 percent, Italy 13.2 percent, United States 10.3 percent and Japan

6.6 percent). The respondents also selected a customer relationship for which they were responsible and with which they had personal experience. Questions about the focal customer relationships appeared during the personal interviews, following a standardized, structured questionnaire with a five-point, Likert-type scale ranging from "strongly disagree" to "strongly agree" or "not at all" to "very much."

Fourth, the database must have sufficient observations to allow for meaningful statistical analysis, including observations for both measure purification (i.e., incremental innovation generation) and causal modeling using multivariate methods (which is beyond the scope of this study). In general, multivariate data analysis demands at least 100 observations and 10 observations for each independent variable (Hair et al. 1998). The IMP database data include more than 200 companies.

Theory and Hypotheses

In this step, the researcher formulates theory-based hypotheses to explore in the case study (Step 4) and refine in Step 5 while classifying the items in the construct. According to Step 2, the "Adaptations" section in the IMP questionnaire likely reflects some dimensions of the innovation generation construct. However, Step 1 indicates that researchers generally favor a discussion of both radical and incremental innovation. Accordingly, in Step 3, the researcher revisits theory to articulate hypotheses about both incremental and radical innovation in buyer–seller relationships. For this purpose, this study uses the S-curve perspective of innovation; S-curves represent a popular and enduring means to study radical and incremental innovation theoretically (Sood and Tellis 2005; Schilling and Esmundo 2009).

Each S-curve represents a dominant design (Abernathy and Utterback 1978) or technology platform (Tushman, Anderson and O'Reilly 1997). Generally, innovation scholars (e.g., Sood and Tellis 2005) agree that innovation progresses along an S-curve as the personnel working with the technology gain facility with it over time and develop more expertise. These personnel may include suppliers who develop a high familiarity with customer processes and machinery and therefore become part of a team, which enhances the incremental innovation on its particular S-curve (Sivakumar and Roy 2004). However, a new technology platform and S-curve also requires new sets of knowledge, both among company personnel and from suppliers (Bower and Christensen 1995; Sivakumar and Roy 2004). New suppliers that are unfamiliar with the customers' organization, process and machinery offer new perspectives on a new S-curve. For example, Christensen (1997) notes that the entire disk drive industry was replaced as new suppliers entered the industry with initially less efficient but smaller hard drives; old suppliers could not test the new versions because their efficiency might suffer. Thus, existing suppliers became extinct as radical innovations by new suppliers appeared. Stated formally:

H1: Incremental innovation occurs in the form of an S-curve within existing buyer–seller relationships on a particular technology platform.

H2: Radical innovation involves new buyer–seller relationships on a new S-curve that initially appears less efficient or more costly than older platforms.

Using formal hypotheses such as these, which are based on the initial database characteristics and theory, the researcher identifies a case study firm. This choice follows theory-building approaches for case studies (Eisenhardt 1989; Yin 2008), such that, for this study the chosen firm should have experienced innovation generation in buyer–seller relationships over time.

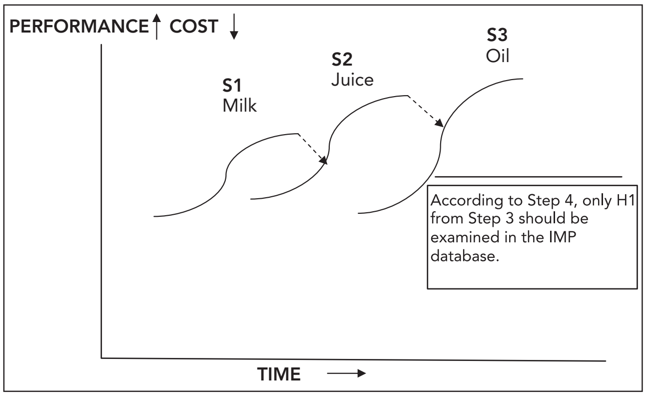
Conducting a Case Study, Exploring Construct Dimensions: An Illustration

For this step, the researcher needs access to a firm that will share information about the phenomena of interest, such as innovation generation in buyer–seller relationships. The case study focuses on the phenomena (Yin 2008) and does not need to apply to a firm in the database. This step works to clarify how the phenomena of interest, as predicted by theory, operate in the real world. By clearly framing the context of the construct, the researcher attempts to match the questionnaire items to the in-depth case study for use in Phase II.

The venue of the case study for this investigation is the Hindustan Packaging Co. Ltd. (HPCL), an Indian joint venture between Tetra Pak, Sweden and the National Dairy Development Board (NDDB), India. The NDDB (http://www.nddb.org) is a statutory body of the Indian government, established in 1964 to spread the Anand pattern of dairy cooperatives throughout India under the auspices of a dairy development program called "Operation Flood." The family-owned Tetra Pak (http://www.tetrapak.com; personal correspondence 1998) is the world leader in aseptic packaging of liquid foods with 1997 sales of US$7.2 billion, with 18,000 worldwide employees. This case study explores three Tetra Pak products developed by HPCL and depicts them as three S-curves that demonstrate collaboration with suppliers and packaging customers in India. The three innovations are the Tetra Hedron for milk (S1), Tetra Brik for juices (S2) and Tetra Brik (S3) for edible oil (see Figure 2).

The in-depth interviews with past and present technical personnel of the joint venture took place during December 1997 and January 1998. Interviewees included the managing director and two past managing directors of HPCL, as well as the chair and founder of NDDB. Technical personnel involved in manufacturing the packaging material, servicing filling machines, supply management, production planning and marketing also participated in interviews. To enhance these interview data, the case study obtained input from archival materials, such as old Tetra

FIGURE 2
S-Curves and Innovation Theory



Note: S1 is the nonaluminum foil Tetra Hedron, S2 is the Tetra Brik juice packaging material and S3 is the Tetra Brik for edible oil.

Pak joint venture files, and published texts (Kamath 1989) dealing with aspects pertinent to the case.

S1: The Base Point: Nonaluminum Foil Tetra Hedron Packaging Material (1982–1985). In 1982, the joint venture developed Tetra Hedron milk packages without aluminum foil and with a 15-day shelf life. This packaging allowed for simple printing that was acceptable to the plain milk market. The pyramid-shaped Tetra Hedron package has no creases, and the base paper is relatively basic paperboard with only a layer of polyethylene applied inside the printed paperboard. This layer makes contact with the milk and seals the package. The technology transfer from Tetra Pak and the training of Indian technicians followed curve S1 in Figure 2.

The HPCL production engineers manufactured the simplest kind of laminate, gradually reducing waste, improving the quality of the raw paperboard and flexographic printing and learning techniques for manufacturing laminated paper. Various supplier development activities resulted in movement up the S1 curve, including the partial use of indigenous bamboo pulp instead of imported pine pulp at the paper supplier's mill, as well as reduction of the extruder neck-in, which was based on work with a U.S. equipment supplier and the plastic granule supplier. The progression up S1 appeared to result from various modifications after the suppliers sent the initial materials.

S2: Radical Innovation in Juice Packaging Material for Tetra Brik (1985–1988). Soon thereafter, Tetra Pak, Sweden, started marketing Tetra Brik packaging systems in India. To run these machines, packaging material containing a layer of aluminum foil needed to be imported; they targeted primarily the soft drink/fruit beverage market. The imported packaging material was very expensive due to high import duties. One product that used Tetra Brik technology, "Frooti" — a mango drink launched by Parle (now the principal Coca-Cola bottler in India) — became an outstanding success. Packaging material came initially from a Singapore plant but then began to be developed at the Indian plant. Parle wanted to buy Indian packaging material, which would be cheaper, and worked with HPCL on a "will pay if it works" basis. After HPCL agreed to make a trial run of a few printed packaging material reels, it rushed the results to Parle's Bombay plant, and within just a few trials, the material started to work. This major innovation for the HPCL plant used manufacturing equipment that was not designed for such manufacturing to create the packaging material. Instead of coating the material with plastic in one pass, the process called for repeat passes that would coat the printed side of the paper with plastic, stick the aluminum foil to the plastic, and then finally place two layers of plastic on the product side. Thus, instead of passing the paper just once through the extrusion line, the new method needed four passes and thus very careful monitoring of the manufacturing process and parameters. To use the existing machinery to produce a radically different packaging material, major developmental work was necessary to improve the upstream suppliers' products, including raw paper that could withstand creasing, splicing tape and modifications to the extrusion coating line and filling machines. With this developmental work, HPCL soon exceeded 100 million packs per year of fruit drink sales, and the mango drink category in India became the largest in the world for the Tetra Pak system.

Between 1985 and 1988, HPCL engineers created a radical innovation, as S2 depicts. This innovation involved an entirely new kind of base paper, finding aluminum foil at a thinness never manufactured in India before, understanding and using ionomer resins (e.g., Surlyn from DuPont), and coaxing a relatively primitive version of an extrusion coating line to produce aluminum foil laminate. The service technicians also learned to deal with the next generation of filling machines, which used the aluminum foil laminate. The classical notion of S-curves maintains that a new S-curve destroys earlier competence, primarily because of the involvement of new suppliers and the discarding of old suppliers with old technology (Christensen 1997). Yet S2 suggests that most employees were retrained and learned the new technology. After S2 initiated, the production of a laminate package for fruit juice required various supplier-related developmental activities, including the development of nylon, instead of rubber, printing plates to ensure the package printing quality met the standards of the juice market and adding aluminum foil suppliers to the existing paperboard supply chain.

S3: Radical Innovation: Edible Oil Packaging Material Innovation in Tetra Brik (1988). In 1988, many 1-L machines imported for milk production were idle because of the slow aseptic milk market, but there was an urgent need to stabilize price fluctuations in India's huge (6 million metric tons) edible oil market. The vast majority (99.97 percent) of the market focused on loose oil, which was prone to the vagaries of commodity trading and adulteration. To allow consumers a more direct say in the oil market, an obvious solution would be packaging oil in Tetra Brik cartons. Soon after its launch, Dhara became the world's largest consumer packaged edible oil brand, mainly packaged in Tetra Brik cartons.

Edible oil creates a major packaging problem: It tends to ooze through normal polyethylene, making the packs greasy. An additional layer of linear low-density polyethylene (LLDPE) can solve this problem, and the Tetra Pak technology used special extruders that would melt LLDPE granules. The joint venture already had added extruders to handle ionomers, like Surlyn, but it still did not have the capability to extrude LLDPE.

The first shipment of Tetra Pak-Italy packaging material to the oil plant in New Delhi revealed packs that were not sealing properly, but the high-profile launch of Dhara was already underway. Therefore, HPCL had literally weeks to create LLPDE packaging material. Conversations with industry peers and an LLDPE film supplier suggested a solution: Instead of melting the LLDPE granules at the packaging plant, it would develop and purchase LLDPE film that could simply be added to the structure, similar to another layer of aluminum foil.

From 1985 to 1988, before the radical innovation S3, both equipment and skills for manufacturing Tetra Brik packaging material and using Tetra Brik filling machines had expanded. That is, increased upward movement occurred on S2 even before the opportunity to create S3 appeared. Interviews confirmed that without experience in foil lamination during the S2, the technicians and suppliers could not have moved to S3 so quickly.

New skills, such as the ability to deal with a plastic film and test oil packages, also were necessary. Working with the juice system (S2) provided HPCL technicians with the necessary competence and confidence and enabled them to leap to S3. After S3 initiated, suppliers again reappeared to help improve performance and reduce costs. For example, the film supplier tried to develop rolls of film that would be similar to the aluminum foil rolls already in use and that could unwind in a similar fashion.

Conclusions from the Case Study and Implications for Construct Pool Items

Packaging industry professionals, aware of the capabilities of the extrusion line, considered these innovations radical, with different S-curves. If nonfoil milk material was S1, aluminum foil juice material was S2, and edible oil material became S3. When a particular S-curve started, it did not necessarily destroy the competencies of an earlier S-curve for the technicians or suppliers. At the end of the first year of introduction of the laminate material for edible oil packaging, all three S-curves persisted. Incremental innovations progressed along all these S-curves, such as adjustments to the thicknesses of the plastic and paper to reduce costs, improvements to set-up and changeover times and streamlining transport contracts so that the same transporter picked up raw paper and delivered finished packaging material to client packaging stations.

Each S-curve also involved some unique suppliers. Thus S3 needed a LLDPE film that only small plastic film manufacturers could provide, whereas S2 needed the patented Surlyn ionomer that could be supplied only by the multinational firm DuPont. These relationships demanded specific innovations to progress along their S-curves, and it became apparent that incremental innovations in the old relationships would involve movement up the S-curve.

The HPCL buyers and development engineers, when asked about developmental activities with each supplier on each S-curve, reported various cost-reducing efforts, such as working with paper suppliers to decrease paper thickness without reducing its properties or with ink suppliers to reduce ink quantity without changing color depth. On the supply side, service engineers recalled many trials undertaken with customer technicians and rapid feedback to the packaging material plant to repeat production and work with suppliers to correct problems. For example, to resolve paper cracks at the filling stage, they kept the raw paper at higher humidity levels, from the paper mill (supplier) to the converting plant (HPCL) and finally to the client (Parle).

In summary, this case study sheds some important light on H1 and H2. In the paper-converting industry, the three S-curves described represent radical innovations that attained success with new suppliers, such as aluminum foil and Surlyn suppliers or LLDP film producers for edible oil. At the next point in the supply chain, HPCL's customers included three different industries and types of organizations: S1 featured the dairy industry, S2 occurred in the juice and soft drink industry and S3 involved the edible oil industry.

In response to the IMP questionnaire items, the HPCL managers agreed that within the context of their supplier or customer relationships, the "Adaptations" section referred to changes on a particular S-curve for both sides. That is, respondents noted changes they had made, which represented incremental innovations on a particular S-curve. The case study participants validated the items; the HPCL buyers provided examples and instances of how suppliers of paper, plastic and ink helped the innovation generation for a particular S-curve. In addition, the service engineers provided instances and examples of how they facilitated adaptation and innovation generation with the customer's packaging plant. For example, HPCL service engineers remained stationed for several weeks at the packaging facility, just to ensure trouble-free operations. Respondents at HPCL also identified the items in the IMP questionnaire that offered the most comprehensive view of their incremental innovation.

This finding suggests the need to drop the radical innovation hypothesis (H2 in Step 3) from the next phase. Although this finding seems somewhat disappointing, the analysis of the incremental innovation data may result in a more significant contribution, considering the robustness of the IMP database and the available data points.

Until this point in Phase I, the researcher has tried to examine the preliminary match among the theory, the domain of interest, and the database fields (in Step 1) and considered the database characteristics and structure (Step 2); articulated theory and preliminary hypotheses for a case study (Step 3); conducted a case study and tested the theory and database item links; and finally revisited the hypotheses (Step 4). In Phase II, the researcher turns to dealing with the data in the database.

PHASE II: MEASURE DEVELOPMENT AND PURIFICATION

Classify Construct Items

Phase II involves the researcher dealing with the numerical data in the survey database. The theory and case study from Steps 1 to 4 inform this step. For the illustration herein, Phase II includes only the incremental innovation generation hypothesis (H1 in Step 3), which pertains to changes in the product itself and its associated processes, such as delivery and information exchange. Therefore, greater innovation involves more adaptations and changes from existing or previous situations. The questionnaire items reflect changes suggested by the customer to adapt to the suppliers' product, as well as the changes that suppliers make to adapt to customer requirements, such as in the product, production process, production schedules, technical information, quality control procedures and so forth.

According to the case study, the items in the IMP questionnaire referring to technical advice from suppliers should appear together with items referring to changes and adaptations made by suppliers. An initial pool of items from the questionnaire appears in Table II; the items relate to changes made by the supplier for customers and changes made by the customer for the supplier.

This process also requires that the researcher verify that the basic descriptive data (e.g., number of firms) in the data set match prior published work available from IMP (Blankenburg-Holm et al. 1996, 1999; Wiley et al. 2006) and address and report any discrepancies. In contrast to one-off surveys, the survey database approach allows for both replication and extension of theory, assuming the researcher provides a clear explanation in this step.

At the end of Step 5, the researcher should have a data set that appropriately labels the innovation generation items and consolidates them using a statistical package such as SPSS or SAS.

Exploratory Factor Analysis

Steps 6 and 7 involve the same data analysis processes for construct measurement as in one-off surveys (Hair et al. 1998), as a prelude to multivariate analysis. Step 6 requires refining the construct statistically (Aquino, Grover, Bradfield and Allen 1999); as the pool of items in Table II indicates, several items from the IMP database may indicate the construct of incremental innovation generation.

The basic data analysis involves testing for skewness and determining the range, mean, mode and standard deviation, all of which are acceptable. Two items have high missing values and are deleted. The EFA relies on SPSS, and a principal axis factoring uses Varimax rotation; these steps suggest deleting two items due to their low communalities.

Three factors thus appear to constitute the innovation generation construct in buyer–seller relationships: technical advice provided by the supplier (IGCHSTEC), technical and product development undertaken by the supplier (IGCHSDEV) and technical development undertaken by the customer (IGCHCUST). Table III summarizes the features of these three factors.

The method computes Cronbach's αs for the three dimensions of innovation generation (Churchill 1979; Grapentine 1995). Averaging the scores of the component items suggests three summated scales, again labeled technical advice provided by the supplier (IGCHSTEC), technical and product development undertaken by the supplier (IGCHSDEV) and technical development undertaken by the customer (IGCHCUST), according to the factor loading. The correlations across the three constructs establish the discriminant validity of the measures. Because the correlations are not high (< 0.6; see Hair et al. 1998), the constructs appear conceptually distinct and different from one another.

Step 6 thus reveals that incremental innovation generation, as captured by IMP data, consists of three dimensions. However, the case study implies that the incremental innovation generation construct represents overall movement along a particular S-curve. To examine unobserved incremental innovation generation in the

TABLE II

Innovation Generation: Pool of Items for Refining

IMP Database Code	Question	Scale
CHCFINP CHCPPROC CHCPSCHE CHCTEXCH CHCQUALC	Changes your *customer* has made in order to adapt to yourselves or to your product regarding: — his final product? — his production processes (techniques)? — his production schedules? — the technical information exchange (documentation, training, staff assignment, etc.)? — his quality-control procedures?	1=none 2=small, e.g., only minor investments in the relationship are made 3=neither small nor large, e.g., some investments are made 4=large, e.g., large changes or investments are made in the relationship 5=very large
CHANGE_C (Reverse coded)	To what extent are the changes the customer has made unique to you or similar to those he has made with regard to other suppliers?	1=unique 2=rather uncommon 3=neither nor 4=rather common 5=standard procedure
CHS_PMOD CHS_PDEV CHS_PCAP CHSPPROC CHSSTOCK CHS_DEL CHS_TADV CHS_MAINT CHS_TECI CHSQUALC	Changes your *company* (or by the intermediary on your behalf) to adapt to the customer or its products or procedures, regarding: — product modification? — new product development for this customer's sake? — your production capacity? — your production processes? — location of your own stocks? — your own delivery procedures? — your technical advisory service? — your maintenance service? — your technical information? — your quality control procedures?	1=none 2=small, e.g., only minor investments in the relationship are made 3=neither small nor large, e.g., some investments are made 4=large 5=very large, e.g., the relationship can only be discontinued at large costs
CHANGE_S (Reverse coded)	To what extent are the changes the customer has made unique to you or similar to those you have made with regard to other suppliers?	1=unique 2=rather uncommon 3=neither nor 4=rather common 5=standard procedure

IMP data that might reflect the three components from Step 6, a CFA helps confirm the findings of the EFA; the second-order CFA model for innovation generation consists of the three first-order factors.

Confirmatory Factor Analysis

The data pertaining to the observed indicators for the three components of innovation generation, IGCHSTEC, IGCHSDEV and IGCHCUST, were analyzed using LISREL. Table IV contains the standardized λ values and fit measures.

The λ values of the standardized solution support the three-factor, second-order model of innovation generation (Brown 2006). The CFA model achieves moderate fit measures, as per the "close to" indices suggested by Hu and Bentler (1999). These include the standardized root mean square residual (SRMR=0.8; Brown 2006), root mean square error of approximation (RMSEA=0.11, which is mediocre according to MacCallum, Browne and Sugawara 1996) and the comparative fit index (CFI=0.90, which is acceptable according to Bentler 1990). Alternative structures (e.g., deleting IGCHCUST) do not improve the model fit.

TABLE III

Innovation Generation: Missing Values, Skewness and Communalities

Item Description	Missing	Communalities	Factor Loading
Technical advisory services by supplier		IGCHSTEC	
— Technical advisory service (CHS_TADV)	3	0.863	81
— Maintenance service (CHS_MAINT)	39	0.794	76
— Technical information (CHS_TECI)	12	0.872	91
Cronbach's α	0.83		
Changes made by *supplier* to adapt to customer product, procedures:		IGCHSDEV	
— Product modification (CHS_PMOD)	3	0.677	79
— New product development (CHS_PDEV)	2	0.718	80
— Production capacity (CHS_PCAP)	3	0.652	52
— Production process (CHSPPROC)	3	0.634	59
— Quality control procedures (CHSQUALC)	42	0.527	53
— Location of stocks (CHSSTOCK)[a]	6	0.280	
— Delivery procedures (CHS_DEL)[a]	2	0.391	
Cronbach's α	0.84		
Changes made by *Customer* to adapt to suppliers' product:		IGCHCUST	
— The final product (CHC_FINP)	14	0.511	51
— Production process (CHCPPROC)	16	0.613	70
— Production schedules (CHCPSCHE)	23	0.440	58
— Technical information exchange including documentation, staff training (CHCTEXCH)	17	0.497	40
— Quality control (CHCQUALC)	59	0.492	
Cronbach's α	0.75		
Unique Changes			
Unique changes made by the customer compared with changes made for other supplier[b]	78	0.350	Deleted
Unique changes made by the supplier compared with changes made for other customer[b]	72	0.414	Deleted

[a]Low communalities deleted.
[b]High missing values deleted.

The second-order CFA suggests that innovation generation has a structure of three first-order factors from the database, reflected in the indicators in Table II. The innovation generation second-order construct increases on the S-curve for incremental innovation, as predicted by theory and the case study.

DISCUSSION

A Checklist for Applying the Proposed Methodology

The proposed method can ensure efficient research that is both impactful and credible and allows many researchers to work with the same data. However, several requirements are necessary to enable researchers to apply the proposed steps in their own research.

The first requirement is access to a credible survey database (e.g., ISM Report on Business, IMP). Credibility can be measured on the basis of the previous academic publications that use the database. In addition, it is useful to have details about the data collection methods and types of respondents. Finally, the effort should be worthwhile if there are more than one cross-sectional sets of data available. For example, if the survey was administered in different regions, like Europe and China for IMP, or is repeated, like the ISM Report on Business, the multiple resulting data sets can allow for geographical comparisons (IMP) or econometric time series modeling (ISM), which should enhance the research opportunities.

The second requirement demands that the researcher find some initial items in the database that suggest the dependent variable (e.g., innovation generation) and independent variables (e.g., trust in buyer–seller relationships). If the data are not sufficiently rich, subsequent steps, including the case study, should not be undertaken.

TABLE IV

Completely Standardized Solution λ: Three-Factor, Second-Order CFA

	Technical Advisory Services by Supplier (IGCHSTEC)	Changes Made by *Supplier* to Adapt to Customer Product Procedures (IGCHSDEV)	Changes Made by *Customer* to Adapt to Suppliers' Product (IGCHCUST)
— Technical advisory service (CHS_TADV)	0.79		
— Maintenance service (CHS_MAINT)	0.75		
— Technical information (CHS_TECI)	0.83		
— Product modification (CHS_PMOD)		0.74	
— New product development (CHS_PDEV)		0.79	
— Production capacity (CHS_PCAP)		0.68	
— Production process (CHSPPROC)		0.68	
— Quality control procedures (CHSQUALC)		0.64	
— The final product (CHC_FINP)			0.53
— Quality control procedures (CHSQUALC)			0.68
— Production schedules (CHCPSCHE)			0.72
— Technical information exchange (CHCTEXCH)			0.53
— Quality control (CHCQUALC)			0.58
Goodness-of-fit statistics			
Chi-square		241	
Degrees of freedom		62	
Standardized root mean square residual (SRMR)		0.082	
Root mean square error of approximation (RMSEA)		0.11	
Comparative fit index (CFI)		0.90	

The researcher also should expend considerable effort and time on theory and hypothesis development before launching the case study. This requirement may be the most critical element of the proposed method, because the more the researcher understands existing theory about the constructs, the better the outcome of the case study and database analysis will be. Theoretical investigation should lead to testable hypotheses for the case study.

The fourth requirement is to locate an appropriate organization for the case study. The organization does not need to be one of the database firms, but it should be qualified on two counts. First, the phenomena of interest should have occurred (e.g., innovation generation in buyer–seller relationships). Second, the firm should offer access to the organization, its personnel and its archives. Organizations that have provided academic access in the past might be identified from the researcher's contacts and frequently will allow access again. Because the researcher has a fairly clear idea about the theory, hypotheses and potentially relevant items in the questionnaire at this point, it may be useful to create a one-page request letter that outlines the research and seeks a meeting with, preferably, the CEO or a designated manager. Case study respondents often express concerns about the excessive time required and confidentiality issues; at this point, the researcher should try to keep any meetings brief and exploit his or her focused theoretical work to seek responses to these precise questions. It also can be helpful to study publicly available company materials (e.g., corporate Web site, annual reports, press releases, trade magazines, industry reports) to facilitate efficient responses during the interviews. At all times, the researcher should remain sensitive to confidentiality issues, respect and protect the identities of respondents and keep firms anonymous if required. Finally, it may be necessary to conduct more than one in-depth case study to flesh out the context and meaning of the database items.

The Phase II steps are similar to those in one-off surveys and should follow the good statistical practices established by previous literature. A potential concern emerges if a construct appears to be inadequately measured in the database; however, this problem can also occur in one-off surveys.

Any researcher conducting database research deals with public databases that have or will produce published research over time. Therefore, researchers have a special responsibility to reconcile all statements made with previous work. If a previous paper using the same database mentions 200 respondents but the current research includes 100, the researcher needs to offer an appropriate clarification. The respondents may have been reduced through data cleaning or the exclusion of data points due to missing data. In this sense, using public databases improves the transparency of the research process and strengthens the claims of the researcher, provided he or she follows the Phase II steps properly and confirms them against previous publications.

Contributions to Theory and Practice

Business researchers thus far have attempted to bridge the rigor–relevance and quantitative–qualitative dilemmas using a dichotomous, either–or approach. For example, as case studies have become more rigorous and quantitative, better questionnaire designs and pretests have attempted to improve the relevance and qualitative strength of large survey databases. Triangulation also serves as a means to study the same phenomena qualitatively and quantitatively while looking for converging evidence. This study suggests a new approach that entails a progression from theory to case study and the database. This approach attempts to build both relevance and rigor into the high-growth, high-opportunity field of database-oriented academic research.

Supply chain research often uses surveys of supply managers, but the logistics and resources of the researcher and the inability of others to replicate and extend the work can limit one-off surveys. In turn, theory development in the supply chain domain becomes more difficult, especially considering the global and dynamic nature of supply chains. The proposed methodology instead allows researchers to use supply chain survey databases (as in the ISM Report on Business, see Kauffman 1999; Cho and Ogwang 2007) to build theory in the supply chain area.

Specifically, this study offers an outline and rationale for a seven-step approach and illustrates the proposed method according to an example of innovation generation in buyer–seller relationships. By providing a checklist, this study also gives researchers a means to evaluate the suitability of the method for their research question. Thus, this article contributes significantly to supply chain research methodology.

REFERENCES

Abernathy, W.J. and J.M. Utterback. "Patterns of Industrial Innovation," *Technology Review*, (80:7), 1978, pp. 40-47.

Aquino, K., S.L. Grover, M. Bradfield and D.G. Allen. "The Effects of Negative Affectivity, Hierarchical Status, and Self-Determination on Workplace Victimization," *Academy of Management Journal*, (42:3), 1999, pp. 260-272.

Bentler, P.M. "Comparative Fit Indices in Structural Models," *Psychological Bulletin*, (107:3), 1990, pp. 238-246.

Blankenburg-Holm, D., K. Eriksson and J. Johanson. "Business Networks and Cooperation in International Business Relationships," *Journal of International Business Studies*, (27:5), 1996, pp. 1033-1053.

Blankenburg-Holm, D., K. Eriksson and J. Johanson. "Creating Value Through Mutual Commitment to Business Network Relationships," *Strategic Management Journal*, (20:5), 1999, pp. 467-486.

Bower, J.L. and C.M. Christensen. "Disruptive Technologies: Catching the Wave," *Harvard Business Review*, (73:1), 1995, pp. 43-53.

Brown, T.A. *Confirmatory Factor Analysis for Applied Research*, The Guildford Press, New York, 2006.

Cho, D.I. and T. Ogwang. "A Conceptual Framework for Computing U.S. Non-manufacturing PMI Indexes," *Journal of Supply Chain Management*, (43:3), 2007, pp. 43-52.

Christensen, C. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Cambridge, MA, 1997.

Churchill, G.J. "A Paradigm for Developing better Measures of Marketing Constructs," *Journal of Marketing Research*, (16:1), 1979, pp. 64-73.

Colgate, M. "Creating Sustainable Competitive Advantage through Marketing Information System Technology: A Triangulation Methodology within the Banking Industry," *The International Journal of Bank Marketing*, (16:2), 1998, pp. 80-89.

Eisenhardt, K.M. "Building Theories from Case Study Research," *Academy of Management Review*, (14:4), 1989, pp. 532-550.

Gerbing, D.W. and J.C. Anderson. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, (25:2), 1988, pp. 186-193.

Ghauri, P. and K. Gronhaug. *Research Methods in Business Studies*, Pearson Education Limited, Essex, England, 2005.

Grapentine, T. "Dimensions of an Attribute: Summated Scales Measure the Relationship between Product Attributes and Their Perceptual Dimensions," *Marketing Research*, (7:3), 1995, pp. 19-27.

Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black. *Multivariate Data Analysis*, Prentice Hall, Upper Saddle River, NJ, 1998.

Hakansson, H. and IMP Project Group. *International Marketing and Purchasing of Industrial Goods*, John Wiley and Sons, Chichester, UK, 1982.

Houghton, J.D. and C.P. Neck. "The Revised Self-Leadership Questionnaire: Testing a Hierarchical Factor Structure for Self-Leadership," *Journal of Managerial Psychology*, (17:7/8), 2002, pp. 672-691.

Hu, L. and P.M. Bentler. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," *Structural Equation Modeling*, (6:1), 1999, pp. 1-55.

Hurley, A.E., T.A. Scandura, C.A. Schriesheim, M.T. Brannick, A. Seers, R.J. Vandenberg and L.J. Williams. "Exploratory and Confirmatory Factor Analysis: Guidelines, Issues, and Alternatives," *Journal of Organizational Behavior*, (18:6), 1997, pp. 667-683.

Kamath, M.V. *Management Kurien Style: The Story of the White Revolution*, Konarak Publishers, New Delhi, India, 1989.

Kauffman, R.G. "Indicator Qualities of the NAPM Report on Business," *Journal of Supply Chain Management*, (35:2), 1999, pp. 29-37.

Kwon, I.W.G. and T. Suh. "Factors Affecting the Level of Trust and Commitment in Supply Chain Relationships," *Journal of Supply Chain Management*, (40:2), 2004, pp. 4-14.

MacCallum, R.C., M.W. Browne and H.M. Sugawara. "Power Analysis and Determination of Sample Size for Covariance Structure Modeling," *Psychological Methods*, (1:2), 1996, pp. 130-149.

Marshall, C.T. and R.D. Buzzell. "PIMS and the FTC Line-of-Business Data: A Comparison," *Strategic Management Journal*, (11:4), 1990, pp. 269-282.

Peter, J.P. "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research*, (16:1), 1979, pp. 6-17.

Pisano, G. "Profiting from Innovation and the Intellectual Property Revolution," *Research Policy*, (35:8), 2006, pp. 1122-1130.

Popadiuk, S. and W.C. Choo. "Innovation and Knowledge Creation: How Are These Concepts Related?," *International Journal of Information Management*, (26:4), 2006, pp. 302-312.

Roy, S., K. Sivakumar and I.F. Wilkinson. "Innovation Generation in Supply Chain Relationships: A Conceptual Model and Research Propositions," *Journal of the Academy of Marketing Science*, (32:1), 2004, pp. 61-79.

Rutner, S.M. and B.J. Gibson. "The Effects of Survey Collection Methods on Reliability, Response Rate and Costs," *Proceedings of the 2001 Council of Logistics Management*, 2001, pp. 299-316.

Scandura, T.A. and E.A. Williams. "Research Methodology in Management: Current Practices, Trends, and Implications for Future Research," *Academy of Management Journal*, (43:6), 2000, pp. 1248-1264.

Schilling, M.A. and M. Esmundo. "Technology S-Curves in Renewable Energy Alternatives: Analysis and Implications for Industry and Government," *Energy Policy*, (37:5), 2009, pp. 1767-1781.

Sheth, J.N. "Organizational Buying Behavior: Past Performance and Future Expectations," *The Journal of Business and Industrial Marketing*, (11:3/4), 1996, pp. 7-24.

Shi, Z., A.S. Kunnathur and T.S. Ragu-Nathan. "Is Outsourcing Management Competence Dimensions: Instrument Development and Relationship Exploration," *Information and Management*, (42), 2005, pp. 901-919.

Sivakumar, K. and S. Roy. "Knowledge Redundancy in Supply Chains: A Framework," *Supply Chain Management*, (9:3/4), 2004, pp. 241-249.

Sood, A. and G.J. Tellis. "Technological Evolution and Radical Innovation," *Journal of Marketing*, (69:3), 2005, pp. 152-168.

Sorescu, A.B., R.K. Chandy and J.C. Prabhu. "Why Some Acquisitions Do Better Than Others: Product Capital as a Driver of Long-Term Stock Returns," *Journal of Marketing Research*, (44:1), 2007, pp. 57-72.

Tushman, M.L., P.C. Anderson and C. O'Reilly. "Technology Cycles, Innovation Steams, and Ambidextrous Organizations: Organization Renewal through Innovation Streams and Strategic Change." In M.L. Tushman and P. Anderson (Eds.), *Managing Strategic Innovation and Change*, Oxford University Press, New York, 1997, pp. 3-23.

Van Prooijen, J.-W. and W.A. Van Der Kloot. "Confirmatory Analysis of Exploratively Obtained Factor Structures," *Educational and Psychological Measurement*, (61:5), 2001, pp. 777-792.

Wiley, J., I. Wilkinson and L. Young. "The Nature, Role and Impact of Connected Relations: A Comparison of European and Chinese Suppliers' Perspectives," *The Journal of Business and Industrial Marketing*, (21:1), 2006, pp. 3-13.

Yin, R.K. *Case Study Research: Design and Methods (Applied Social Research Methods)*, Sage Inc., Thousand Oaks, CA, 2008.

Subroto Roy (Ph.D., University of Western Sydney) is associate professor of marketing at the University of New Haven, West Haven, Connecticut. His primary areas of interest are innovation, business-to-business marketing and supply chains, and his blog can be seen at www.stratoserve.com. Before entering academia, Dr. Roy worked for the Indian Joint Venture of Tetra Pak Sweden, serving as Head of Marketing and Sales among other positions. He has published research in *Industrial Marketing Management, International Journal of Marketing Research, Journal of the Academy of Marketing Science* and *Journal of Business Research*, among other publications. In spring 2011, Dr. Roy will be a Visiting Scholar at Yale University.