Journal of Purchasing & Supply Management ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Purchasing & Supply Management

journal homepage: www.elsevier.com/locate/pursup

Experiments in purchasing and supply management research

Stephanie Eckerd

Department of Marketing and Supply Chain Management, James A. Haslam II College of Business, University of Tennessee, 916 Volunteer Boulevard, 325 Stokely Management Center, Knoxville, TN, 37996 United States

ARTICLE INFO

Article history:
Received 28 July 2016
Accepted 22 August 2016

Keywords:
Supply chain management research
Experiments

ABSTRACT

In this article, I provide a brief introduction to the use of experiments in purchasing and supply management research. Different types of experiments common within the field are introduced, and within these types I compare the strengths and weaknesses regarding reliability and validity concerns and the limitations surrounding participant pools. Topics for which experimentation is appropriate and useful are identified, and examples of recent work in the field provided. Several trends and future research opportunities are presented to inspire more work in this arena.

1. Introduction to the method

1.1. What are experiments?

Business processes, including those in purchasing and supply management (P/SM), are designed, managed, and continuously improved by people, and are therefore subject to the decision biases and bounded rationality of those people. In many cases, this human intervention with systems and processes can lead to sub-optimal performance. Behavioral experiments are used to investigate the world of factors influencing human judgment and decision-making. Experiments are useful for helping to explain the decisions occurring in practice, as compared to theoretical predictions. Experiments are further useful in designing and testing mechanisms to facilitate better decision-making.

Experiments allow researchers to observe directly the impact of a change in a factor (i.e., the treatment) on an outcome. There are many different types of experiments, ranging from controlled laboratory experiments, to scenario-based experiments, to field experiments.

- *Laboratory-based experiments* in the social sciences assemble participants in a common space, typically a computer laboratory, where they conduct a specific task (Webster and Sell, 2014). In operations and supply chain management, the decisions made in these tasks are often quantifiable, for example how much to order or the setting of a particular contract parameter, and the results can be compared to an analytically derived theoretical optimal. One example of a laboratory experiment conducted in the P/SM space is that of Eckerd et al. (2013), in which they evaluate the effect of psychological contract breaches and associated emotional responses on buyers' order quantities over time.
- *Scenario-based experiments* (sometimes referred to as vignette experiments) present a written (or video graphed) description of a situation or event to participants, and then ask participants to make decisions based on what they have read (Rungtusanatham et al., 2011). These decisions can include quantitative outcomes such as what the participant is willing to pay for something, but often focus on more qualitative outcomes such as which supplier the participant prefers or to what extent the participant trusts a supplier. Facilitated by the nature of the questions, this type of experiment will often be administered on-line and is thus similar to a survey, although scenario-based experiments can, and often are, conducted in a more controlled laboratory setting, as well. An example of a scenario-based experiment in the P/SM space is Thomas et al.'s (2013) evaluation of the impact of different negotiation strategies on information exchange, communication quality, and operational knowledge transfers.
- *Field experiments* are those conducted in an actual work environment (Chatterji et al., 2016). These experiments often assess the change in a process after a treatment is introduced, thus setting up a pre-test post-test comparison of the effect of the treatment. While field experiments are quite rare in the P/SM literature, one such example is provided in Hardgrave et al. (2013). Using longitudinal data collected from two large retailers during the course of two field experiments, they investigate the efficacy of Radio Frequency Identification in reducing inventory inaccuracy.

While these examples are not exhaustive of the types and uses of experiments available, they do represent the types more

E-mail address: seckerd@utk.edu

http://dx.doi.org/10.1016/j.pursup.2016.08.002
1478-4092/

commonly appearing in or applicable to the P/SM literature.

1.2. Why use this method in the P/SM literature?

While the application of experiments within the P/SM literature is rather nascent, topic areas within P/SM are particularly well positioned for study using experiments. This is because so many behavioral aspects (Boudreau et al., 2003) are fundamental to the work conducted in P/SM: the work is largely conducted by people (not machines) and thus people are a major factor; their work is not always perfectly observable; the people and the environments they work in are not deterministic and predictable; supply chains and the actors working within them are highly dependent on one another; the people conducting the work are subject to learning and fatigue and are expected to solve difficult problems; the workers themselves are often part of the product or service being performed; and, these workers have emotions, and despite a strong belief by many that emotions do not enter into the workplace, sufficient research and anecdotal evidence demonstrates otherwise. Research using experiments can help improve our understanding of the manifestations of these behavioral aspects in the world of P/SM, thereby providing the necessary foundation for improving the quality and effectiveness of decisions in the field.

Table 1 highlights recent works (since 2010) focusing within several areas relevant to P/SM. Although the table is by no means an exhaustive list, it is used to illustrate the types of problems that can be studied using experiments, the variety of different approaches to experimentation that have been used in P/SM, and the participant pools that are commonly utilized (further discussed in the Limitations section, below).

2. Benefits of using experiments

Experiments are useful because they afford the researcher an enormous amount of control, are generally very efficient to run, and are replicable (Siemsen, 2011). These aspects of experimentation make it key in testing and refining theory, as well as in facilitating the evaluation of inconsistencies of theories in practice (Tokar, 2010). One of the most widely researched areas within P/SM involving experiment methodology is that of the bullwhip effect. Through carefully designed and executed experiments, researchers were able to narrow in on both the operational and behavioral causes of the phenomenon in a way that other methods in isolation simply would not accomplish (Ancarani et al., 2016; Croson et al., 2014; Croson and Donohue, 2006). This stream of research has developed our theoretical and practical understanding of the bullwhip phenomenon.

Experimentation also provides the ability to evaluate the implementation of new strategies, procedures, and policies in a manner that is affordable and realistic (Croson and Gachter, 2010). Management of supply chain disruptions is but one area within P/SM for which this aspect is particularly attractive. Many supply disruptions (think natural disasters, terrorist attacks) are unpredictable and extraordinarily large in scale. As researchers, we might want to plan and execute data collection efforts following these events, but how can we reasonably engage in such efforts without adding more disruption to an already all-consuming event? Yet, clearly the study of such phenomena is necessary to improve both theoretical development and practice. One way we can accomplish this meaningfully is through experiments designed to capture the critical elements surrounding a disruption; the "safe" replication of an event without the myriad costs associated with it, including the potential cost of human life. This leverages the same ideas as soldiers preparing for training or critical war room exercises conducted at the upper echelons of organizations. Experiments of this nature are thus incredibly useful; however, one trade-off is that they do relinquish some degree of control for a more realistic contextual experience, similar to that of a field experiment. The implications of this trade-off are discussed more in the section below on internal and external validity.

3. Limitations of using experiments

One of the primary criticisms of experiments as a methodology involves the use of students as subjects. The perception is often that students are used in experiments because they are cheap and easy to recruit, whereas the challenge associated with involving managers - bringing them into a laboratory and appropriately compensating them for their time - is too steep. Certainly there is truth under-riding the observation, but the reality is that for many experiments the decision-making behaviors being evaluated are relevant to *people* – in other words they study, by intention, *human* behavior and not necessarily just *manager* behavior. For example, the very hallmarks of experiments testing economic theory (i.e., context-free, randomized) make the specific individuals and their "life baggage" immaterial to that research effort. Of course, this is not to suggest that people, their experiences, the environment, and so on, do not matter or influence decision-making. However, these aspects can be studied using other methodologies, like surveys and case studies, which are then used in triangulation to create a more comprehensive interpretation of the relationships under investigation (Boyer and Swink, 2008). In experiments, the use of students as subjects should be bounded to those research efforts bearing a universalistic conceptualization (Gordon et al., 1987; Stevens, 2011). As the research questions become more particularistic and concrete, as they typically do in scenario-based experiments for example, then a different participant pool is often warranted. This is evidenced in Table 1, where the scenario-based

Table 1
Select P/SM Literature using Experiments.

Topical Area within P/SM	Authors (Year)	Type of Experiment	Type of Participant
Information Availability	Morssinkhof et al. (2011)	Laboratory	Undergraduate students and practitioners
	Haines et al. (2010)	Laboratory	Undergraduate students
Inventory Control	Tokar et al. (2016)	Scenario-based and Laboratory	Undergraduates, MBA students, and managers
	Hardgrave, Aloysius, and Goyal (2013)	Field	Store-level inventory data, employees
Negotiations	Ribbink and Grimm (2014)	Laboratory	MBA students
	Thomas et al. (2013)	Scenario-based	Undergraduate students
Relational and Emotional Aspects	Murfield et al. (2016)	Scenario-based	Supply chain professionals
	Eckerd et al. (2013)	Laboratory	Undergraduate and MBA students
Supplier Selection	Kull et al. (2014)	Scenario-based	Supply chain professionals
	Hada et al. (2013)	Scenario-based	Purchasing managers and VPs
Supply Disruption	Ro et al. (2016)	Scenario-based	Full-time working professionals
	Cantor et al. (2014)	Scenario-based	MTurk (Amazon worker marketplace)

Please cite this article as: Eckerd, S., Experiments in purchasing and supply management research. Journal of Purchasing and Supply Management (2016), http://dx.doi.org/10.1016/j.pursup.2016.08.002

experiments typically involve respondents who are working professionals, as these are the people falling within the boundary conditions of the theory being tested (Thomas, 2011).

On a more personal note, one of the toughest lessons that I learned when first starting out with experiments is that different disciplines have their own idiosyncratic rules regarding "correct" experimental practice. Those trained in experimental economics adhere to induced valuation (paying subjects for performance), and this makes sense given the types of questions they tend to ask. Moreover, deception, a practice that is not atypical within psychology-based experiments and again may make perfect sense given the types of questions they are asking, is a strictly forbidden practice in experimental economics. These are not necessarily limitations, but are nonetheless hotly contested factors that a researcher interested in conducting experiments has to be aware of. These particular aspects of design may complicate and prolong ethics processing with the researcher's institutional review board (IRB), and will almost certainly have an impact on the journal outlets available for the research. Many primers exist to help researchers wade through these philosophical viewpoints (see Croson, 2005, for an insightful introduction to the topic).

4. Reliability and validity in experiments

Experiments are reliable when it is possible to replicate the experiment and generate similar results (Carmines and Zeller, 1979). This generally means that the instruments used in experiment administration and measurement have been shown to be consistent across numerous studies. Reliability in laboratory experiments is improved by using scripted research protocols and the same physical space and proctor in all sessions. Reliability also involves assurance that the participants are comfortable and confident in making the decisions they are asked to make. This is often accomplished through training exercises, and can also involve the use of decision support systems that simplify the decision-making burden. In experiments where the task involves some form of quantifiable decision-making task, tests can be administered to ensure participant understanding; subjects unable to pass the test are excused from the study or their data not included in further analysis.

Experimental validity speaks to whether the test is measuring the causal relationship it purports to measure (i.e., internal validity), and addresses the extent to which the research results are generalizable to the larger population (i.e., external validity) (Mentzer and Flint, 1997). The different types of experiments introduced in Section 1 vary in the degree of internal versus external validity achievable. A strictly controlled laboratory experiment is typically very high on internal validity - if everything else is held constant, we can be reasonably confident that the treatment is causing the change or behavior observed. However, these types of experiments also tend to be quite low on external validity - we usually cannot make statements as to how well the observations hold in a real world setting and what additional influences in the real world would affect the relationship observed. Field experiments often occur at the other extreme, in that it is very difficult to control all externalities when dealing with real people and processes (i.e., lower internal validity) but the situation in which the experiment takes place better mirrors the conditions under which the phenomena might actually be observed (i.e., improved external validity).

Certain checks can be conducted to support the validation of experiment research. Manipulation checks assess whether participants are interpreting the experimental treatments as intended and portrayed by the researchers (Rungtusanatham et al., 2011). When an experiment is used to test multiple factors of interest, then it is also necessary to conduct confound checks to ensure that the factors are indeed independent and not inadvertently influencing one another (Bachrach and Bendoly, 2011). Scenario-based experiments will also typically include realism checks to assess the degree to which the participants took their role seriously and perceived the situation to be a believable one (Rungtusanatham et al., 2011).

5. Future research opportunities using experiments

There are many opportunities for future research using experiments in P/SM. Generally speaking, field experiments are vastly underutilized in this space, yet offer a valuable opportunity for enhancing the external validity of research being conducted in the area. The efficiency of experimentation also affords an opportunity for research to incorporate multiple experiments within a single study. In this way, the research can identify and describe a phenomenon, and also design and demonstrate a managerial prescription for it. Knemeyer and Naylor (2011) point out that the packaging of multiple experiments into a single study has been the trend in marketing for some time, and is likely necessary in order to best understand a relationship. Some of the more recent research efforts in P/SM already show this same progression (for example, see Tokar et al., 2016). Alternatively, it may be that multi-methods studies are used to achieve a more multi-faceted understanding of a phenomenon. This, too, is a trend we are seeing more of in the discipline's journals. For example, research by Niranjan et al. (2011) use laboratory experiments coupled with case study to derive a "correction model" of interpersonal dynamics to explain supply line underweighting in a stock management problem. Studies such as this are useful in that the leveraging of multiple methods can be complementary in terms of the trade-offs discussed previously; one method's weakness is the other's strength.

Experiments can offer a clear contribution and complement to many research efforts. Numerous resources are available to the interested reader on how to conduct experiments, experimental design, and protocol; the references provided throughout this primer offer a good place to start. Certainly, based on the subset of topical areas presented in Table 1, there are few limitations regarding the problem areas that experiments can help to address. What will be important for future research is to build out comprehensive research programs investigating issues from multiple angles that include, but do not rest solely on, experiment methods.

References

Ancarani, A., Di Mauro, C., D'Urso, D., 2016. Measuring overconfidence in inventory management decisions. J. Purch. Supply Manag. 22 (3), 171–180.
Bachrach, D.G., Bendoly, E., 2011. Rigor in behavioral experiments: a basic primer for supply chain management researchers. J. Supply Chain Manag. 47 (3), 5–8.
Boudreau, J., Hopp, W., McClain, J.O., Thomas, L.J., 2003. On the interface between operations and human resources management. Manuf. Serv. Oper. Manag. 5 (3), 179–202.
Boyer, K.K., Swink, M.L., 2008. Empirical elephants – why multiple methods are essential to quality research in operations and supply chain management. J. Oper. Manag. 26 (3), 338–344.
Cantor, D.E., Blackhurst, J.V., Cortes, J.D., 2014. The clock is ticking: the role of uncertainty, regulatory focus, and level of risk on supply chain disruption decision-making behavior. Transp. Res. Part E: Logist. Transp. Rev. 72, 159–172.
Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment, Sage. Publications,, Thousand Oaks, CA.
Chatterji, A.K., Findley, M., Jensen, N.M., Meier, S., Nielson, D., 2016. Field experiments in strategy research. Strateg. Manag. J. 37 (1), 116–132.
Croson, R., 2005. The method of experimental economics. Int. Negot. 10, 131–148.
Croson, R., Donohue, K., 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. Manag. Sci. 52 (3), 323–336.
Croson, R., Donohue, K., Katok, E., Sterman, J., 2014. Order stability in supply chains:

Please cite this article as: Eckerd, S., Experiments in purchasing and supply management research. Journal of Purchasing and Supply Management (2016), http://dx.doi.org/10.1016/j.pursup.2016.08.002

coordination risk and the role of coordination stock. Prod. Oper. Manag. 23 (2), 176–196.
Croson, R., Gachter, S., 2010. The science of experimental economics. J. Econ. Behav. Organ. 73 (1), 122–131.
Eckerd, S., Hill, J.A., Boyer, K.K., Donohue, K., Ward, P., 2013. The relative impact of attribute, severity, and timing of psychological contract breach on behavioral and attitudinal outcomes. J. Oper. Manag. 31 (7–8), 567–578.
Gordon, M.E., Slade, L.A., Schmitt, N., 1987. Student guinea pigs: porcine predictors and particularistic phenomena. Acad. Manag. Rev. 12 (1), 160–163.
Hada, M., Grewal, R., Lilien, G.L., 2013. Purchasing managers' perceived bias in supplier-selected referrals. J. Supply Chain Manag. 49 (4), 81–95.
Haines, R., Hough, J.R., Haines, D., 2010. Individual and environmental impacts on supply chain inventory management: an experimental investigation of information availability and procedural rationality. J. Bus. Logist. 31 (2), 111–128.
Hardgrave, B.C., Aloysius, J.A., Goyal, S., 2013. RFID-Enabled visibility and retail inventory record inaccuracy: experiments in the field. Prod. Oper. Manag. 22 (4), 843–856.
Knemeyer, A.M., Naylor, R.W., 2011. Using behavioral experiments to expand our horizons and deepen our understanding of logistics and supply chain decision-making. J. Bus. Logist. 32 (4), 296–302.
Kull, T.J., Oke, A., Dooley, K.J., 2014. Supplier selection behavior under uncertainty: contextual and cognitive effects on risk perception and choice. Decis. Sci. 45 (3), 467–505.
Mentzer, J.T., Flint, D.J., 1997. Validity in logistics research. J. Bus. Logist. 18 (1), 199–216.
Morssinkhof, S., Wouters, M., Warlop, L., 2011. Effects of providing total cost of ownership information on attribute weights in purchasing decisions. J. Purch. Supply Manag. 17, 132–142.
Murfield, M.L.U., Esper, T.L., Tate, W., Petersen, K.J., 2016. Supplier role conflict: an investigation of its relational implications and impact on supplier accommodation. J. Bus. Logist. 37 (2), 168–184.
Niranjan, T.T., Wagner, S.M., Bode, C., 2011. An alternative theoretical explanation and empirical insights into overordering behavior in supply chains. Decis. Sci. 42 (4), 859–888.
Ribbink, D., Grimm, C.M., 2014. The impact of cultural differences on buyer-supplier negotiations: an experimental study. J. Oper. Manag. 32 (3), 114–126.
Ro, Y.K., Su, H., Chen, Y., 2016. A tale of two perspectives on an impending supply disruption. J. Supply Chain Manag. 52 (1), 3–20.
Rungtusanatham, M., Wallin, C., Eckerd, S., 2011. The vignette in a scenario-based role-playing experiment. J. Supply Chain Manag. 47 (3), 9–16.
Siemsen, E., 2011. The usefulness of behavioral laboratory experiments in supply chain management research. J. Supply Chain Manag. 47 (3), 17–18.
Stevens, C.K., 2011. Questions to consider when selecting student samples. J. Supply Chain Manag. 47 (3), 19–21.
Thomas, R.W., 2011. When student samples make sense in logistics research. J. Bus. Logist. 32 (3), 287–290.
Thomas, S.P., Thomas, R.W., Manrodt, K.B., Rutner, S.M., 2013. An experimental test of negotiation strategy effects on knowledge sharing intentions in buyer-supplier relationships. J. Supply Chain Manag. 49 (2), 96–113.
Tokar, T., 2010. Behavioural research in logistics and supply chain management. Int. J. Logist. Manag. 21 (1), 89–103.
Tokar, T., Aloysius, J., Waller, M., Hawkins, D.L., 2016. Exploring framing effects in inventory control decisions: violations of procedural invariance. Prod. Oper. Manag. 25 (2), 306–329.
Webster, M., Sell, J., 2014. Laboratory Experiments in the Social Sciences, 2nd ed. Academic Press,, London.

Please cite this article as: Eckerd, S., Experiments in purchasing and supply management research. Journal of Purchasing and Supply Management (2016), http://dx.doi.org/10.1016/j.pursup.2016.08.002