Available online at www.sciencedirect.com
ScienceDirect
Journal of Operations Management 25 (2007) 785–805
JOURNAL OF OPERATIONS MANAGEMENT
www.elsevier.com/locate/jom

Defining and developing measures of lean production

Rachna Shah [a,*], Peter T. Ward [b,1]

[a] University of Minnesota, Carlson School of Management, 321 19th Avenue South, Minneapolis, MN 55455, United States
[b] The Ohio State University, Fisher College of Business, 600 Fisher Hall, Columbus, OH 43221, United States

Available online 20 January 2007

Abstract

Our research addresses the confusion and inconsistency associated with "lean production." We attempt to clarify the semantic confusion surrounding lean production by conducting an extensive literature review using a historical evolutionary perspective in tracing its main components. We identify a key set of measurement items by charting the linkages between measurement instruments that have been used to measure its various components from the past literature, and using a rigorous, two-stage empirical method and data from a large set of manufacturing firms, we narrow the list of items selected to represent lean production to 48 items, empirically identifying 10 underlying components. In doing so, we map the operational space corresponding to conceptual space surrounding lean production. Configuration theory provides the theoretical underpinnings and helps to explain the synergistic relationships among its underlying components.

Keywords: Lean production; Scale development; Confirmatory factor analysis

1. Introduction

In 360 BC, Plato (in *Cratylus*) suggested that linguistic confusion arises because multiple terms may refer to the same object or idea, a single term may refer ambiguously to more than one object or idea, and terms may be confusing because they are out of date. The same observations can be made today with respect to a number of management approaches. The current study addresses these issues with regard to lean production. We believe that the price paid for lacking a clear, agreed-upon definition is high because empirical testing of inexact and imprecise concepts lead to a body of research that examines slightly different aspects of the same underlying constructs masked by different terminology. Consequently, results from such testing do not improve our understanding, make marginal contributions to the existing knowledge base, and prevent academic fields from making real progress (Meredith, 1993). If theory and empirical work are to advance in this area, semantic differences between lean production and its predecessors must be resolved, the conceptual definition of lean production must be clarified, and operational measures must be more clearly specified. In this paper, we address these three issues.

The approach now known as lean production has become an integral part of the manufacturing landscape in the United States (U.S.) over the last four decades. Its link with superior performance and its ability to provide competitive advantage is well accepted among academics and practitioners alike (e.g., Krafcik, 1988; MacDuffie,

* Corresponding author. Tel.: +1 612 624 4432; fax: +1 612 624 8804.
E-mail addresses: rshah@csom.umn.edu (R. Shah), ward.1@osu.edu (P.T. Ward).
[1] Tel.: +1 614 292 5294.

0272-6963/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.jom.2007.01.019

1995; Pil and MacDuffie, 1996; Shah and Ward, 2003; Wood et al., 2004). Even its critics note that alternatives to lean production have not found widespread acceptance (Dankbaar, 1997) and admit that "lean production will be the standard manufacturing mode of the 21st century" (Rinehart et al., 1997, p. 2). However, any discussion of lean production with managers, consultants, or academics specializing in the topic quickly points to an absence of common definition of the concept.

This lack of clarity is evident from the multiplicity of descriptions and terms used with respect to lean production. The ambiguity stems in part because lean production evolved over a long period of time (Hopp and Spearman, 2004; Womack et al., 1990; Spear and Bowen, 1999) and because of its mistaken equivalence with other related approaches. Hopp and Spearman (2004) note that using closely related terms in the titles of some of the earliest publications may have also contributed to this confusion (see for example Sugimori et al., 1977). These primarily semantic differences between lean and its predecessors are unfortunate but can be resolved fairly easily. A greater source of confusion, however, is the more substantive disagreement about what comprises lean production and how it can be measured operationally.

Our objectives in this paper are as follows. First, we attempt to resolve the semantic confusion surrounding lean production and explain the different perspectives invoked in describing it using a historical evolutionary lens. Second, in our pursuit of a commonly agreed upon definition of lean production, we propose a conceptual definition that encompasses its underlying multidimensional structure. Finally, using a rigorous empirical method, we identify a set of 48 items to measure lean production and its main components. Additionally, we chart the linkages among the items and the components and map the operational space as it corresponds to the conceptual space. In short, we develop the concept of lean production based on extant knowledge and use data from a sample of manufacturers to develop an operational measure that consists of 10 reliable and valid scales.

2. Historical background

Defining lean production requires first examining its historical evolution and identifying the different perspectives that are commonly invoked in describing it. We highlight the key phases that have contributed to our current understanding of lean production in Table 1. Lean production directly descended from and is frequently used as a proxy for Toyota Production System (TPS), which itself evolved from Taiichi Ohno's experiments and initiatives over three decades at Toyota Motor Company. TPS was formally introduced in the U.S. in 1984 when NUMMI was established as a joint venture between Toyota and General Motors, but its informal transfer to the U.S. began much earlier, occurring over time in a piecemeal fashion. A consequence of the slow geographic dispersion separated by a significant time lag was that the understanding of the new system in the U.S. evolved even more slowly and with an additional time-lag.

Because TPS was multifaceted and complicated, it was not easy for U.S. managers to comprehend the true nature of the production process. As in the age-old fable of the blind men touching different parts of an elephant and imagining very different animals, these managers often focused on a single, visible aspect of the process while missing the invisible, highly inter-dependent links of the system as a whole. By the time U.S. managers realized the numerous elements underlying TPS, and, by extension, lean production, these different terms had become deeply ingrained in the common lexicon of the academic and business publications. As a result, semantic discrepancies crept in even when no substantive difference was apparent. Currently, there are numerous academic and practitioner books and articles, yet we still do not have a precise and agreed upon way of defining or measuring lean production.

2.1. Lean production—a literature review

Reviewing the existing literature provides a starting point in defining lean production. Additionally, it helps us highlight the confusion in the conceptual and the operational space surrounding lean production and glean a set of operational measures that can be used to represent it. In conducting our review, we began with the earliest publications related to Japanese manufacturing/production systems ending with the most recent publications related to lean production. We observed that, in general, the early Japanese books were more precise in defining TPS and in identifying its underlying components (Monden, 1983; Ohno, 1988) compared to the research articles because the latter focused on defining and describing specific components of the system rather than the whole (Sugimori et al., 1977; Monden, 1981b). However, the distinction between the system and its components was missed by most early observers, perhaps because of the articles and monographs related to the components were published before the books (in English) that described the system.

This lack of distinction between the system and its components was further complicated by the general

Table 1
Time line marking the critical phases in the lean production evolution

Period	Developments
1927 and before	• Henry Ford outlines his production philosophy and the basic principles underlying the revolutionary Ford Production System (FPS) in "Today and tomorrow" in 1927.
1945-78 Progress In Japan	• 1937 - Toyoda (later Toyota) Motor Company is established in Koromo, Japan. o Toyoda cousins Kiichiro and Eiji, with Taiichi Ohno study FPS and perfect the principle concepts and tools constituting Toyota Production System (TPS). Just in time (JIT) production method is a key component of TPS. • 1978 - Ohno publishes "Toyota Production System" in Japanese. He credits FPS and the American supermarket behind his just in time thinking. o According to Ohno, the primary goal of TPS is cost reduction (waste elimination); it can be achieved through quantity control, quality assurance, and respect for humanity. He recommends producing only the kind of units needed, at the time needed and in the quantities needed.
1973-88 TPS arrives in North America	• 1973 - Oil crisis hits North America and generates immense interest in the (new) Japanese manufacturing and management practices followed by publication of numerous academic and practitioner books and articles. • 1977 - First academic article is published by Sugimori et al.; Narrowly focused articles on topics such as Kanban and just in time production (Monden, 1981b), production smoothing and level loading (Monden, 1981c) appear. • 1984 - NUMMI, a joint venture between Toyota Motor Company and General Motors opens in California. • Mid 1980s - Noteworthy books including Monden's Toyota Production System (1983); Ohno's Toyota Production System: Beyond large-scale production (1988) are published in English. • There is only a piecemeal understanding of TPS and its constituent elements; equivalence between JIT production, kanban and TPS is suggested (see Table 2).
1988-2000 Academic progress	• 1988 - Krafcik coins the term "lean" to describe the manufacturing system used by Toyota. • 1990 - The machine that changed the world by Womack, Jones and Roos is published. o The machine establishes "lean production" to characterize Toyota's production system including its underlying components in the popular lexicon. o The book describes a lean system in detail; but does not offer a specific definition. • Mid 1990s - Articles related to measuring just in time (Sakakibara et al., 1993; Flynn et al., 1995; McLachlin, 1997), total quality management (Ross, 1993; Dean and Bowen, 1994; Sitkin et al., 1994; Flynn et al., 1995), their interrelationships (Flynn et al., 1995; Sakakibara et al., 1997) and the impact of other organizational variables on their implementation are published in the academic journals. • 1994 - Lean Thinking by Womack and Jones is published. The book extends the philosophy and the guiding principles underlying lean to an enterprise level.
2000-present	• Numerous books and articles written by practitioners and consultants, and a few academic conceptual (Hopp and Spearman, 2004; de Treville and Antonakis, 2006) and empirical articles (Shah and Ward, 2003) highlighting the overarching nature of lean production are published; yet no clear and specific definition is available. • 2006 – Toyota Motor Company is projected to become #1 automobile manufacturer in North America.

point of reference used in its description. Lean production is generally described from two points of view, either from a philosophical perspective related to guiding principles and overarching goals (Womack and Jones, 1996; Spear and Bowen, 1999), or from the practical perspective of a set of management practices, tools, or techniques that can be observed directly (Shah and Ward, 2003; Li et al., 2005). This difference in orientation does not necessarily imply disagreement, but it does undermine conceptual clarity.

Our literature review indicates that there exist many descriptions of lean production and its underlying components along with a few conceptual definitions (Table 2); rather than provide a comprehensive list, our intention is to capture the salient similarities between the terms. We observe that the descriptions/definitions are very general and have become more expansive over time. A case in point is just-in-time (JIT), one of the four main concepts of TPS (Monden, 1983). To maintain just-in-time production in Toyota plants, Ohno (1988) devised kanban as a means to pull material from an upstream station and manage product flow. In describing and measuring JIT, Sugimori et al. (1977) also focused on its most critical components such as kanban, production smoothing, and set up time reduction; later definitions incorporate these elements but also include quality improvement and employee involvement (Hall, 1987; McLachlin, 1997) and customer focus (Flynn et al., 1995a,b). Subsequently in the U.S., JIT became *the* system—TPS, pull production, and kanban assumed equivalence with JIT, and the terms are often used interchangeably (Hopp and Spearman, 2004). We observed similar concept stretching in other components of lean production such as quality management, people management, and preventive maintenance approaches.

We also found a well-developed literature base associated with the operational instruments used to measure the components of lean production and observed a similar overlap and confusion in that literature (Table 3). In discussing the operational

Table 2
Lean production—mapping the conceptual definitions

Illustrative definitions
Toyota Production System (TPS) and lean production
The basic idea in TPS is to produce the kind of units needed, at the time needed and in the quantities needed such that unnecessary intermediate and finished product inventories can be eliminated. Three sub-goals to achieve the primary goal of cost reduction (waste elimination) are quantity control, quality assurance, and respect for humanity. These are achieved through four main concepts: JIT, autonomation, flexible workforce, and capitalizing on worker suggestion and 8 additional systems (Monden, 1983, p. 2)
The basis of TPS is the absolute elimination of waste. The two pillars needed to support the TPS are JIT and autonomation (Ohno, 1988)
TPS can be described as an effort to make goods as much as possible in a continuous flow (Ohno, 1988)
Lean production uses half the human effort in the factory, half the manufacturing space, half the investment in tools, half the engineering hours to develop a new product in half the time. It requires keeping half the needed inventory, results in many fewer defects, and produces a greater and ever growing variety of products (Womack et al., 1990, p. 13)
TPS includes standardization of work, uninterrupted work flows, direct links between suppliers and customers, and continuous improvement based on the scientific method (Spear and Bowen, 1999)
Lean production is an integrated system that accomplishes production of goods/services with minimal buffering costs (Hopp and Spearman, 2004)
Just in time (JIT)
Just in time production system as ‘‘only the necessary products, at the necessary time, in the necessary quantity’’ (Sugimori et al., 1977)
Kanban system, production smoothing and setup time reduction are critical components of any JIT system (Monden, 1981b)
JIT philosophy is associated with three constructs: total quality, people involvement, and JIT manufacturing techniques (Hall, 1987)
Programs associated with JIT include ‘‘elimination of waste, and full utilization of people, equipment, materials, and parts’’ (Davy et al., 1992)
JIT is a comprehensive approach to continuous manufacturing improvement based on the notion of eliminating all waste in the manufacturing process (Sakakibara et al., 1993)
JIT is based on the notion of eliminating waste through simplification of manufacturing processes such as elimination of excess inventories and overly large lot sizes, which cause unnecessarily long customer cycle times (Flynn et al., 1995a,b)
JIT is composed of three overall components, namely, flow, quality and employee involvement (McLachlin, 1997)
Total quality management (TQM)
TQM is an integrated management philosophy and set of practices that emphasizes continuous improvement, meeting customer requirements, reducing rework, long range thinking, increased employee involvement and teamwork, process redesign, competitive benchmarking, team-based problem solving, constant measurement of results, and closer relationships with suppliers (Ross, 1993)
TQM is a philosophy or an approach to management that can be characterized by its principles, practices and techniques. Its three principles are customer focus, continuous improvement, and teamwork (Dean and Bowen, 1994)
Common guiding TQM precepts can be conceptually distinguished into three clusters (a) focusing on customer satisfaction (b) stressing continuous improvement, and (c) treating the organization as a total system (Sitkin et al., 1994)
TQM is an approach to improving the quality of goods and services through continuous improvement of all process, customer driven quality, production without defects, focus on improvement of processes rather than criticism of people and data driven decision making (Flynn et al., 1994)

Table 3
Lean production—charting the measurement instruments

Scale/individual measure [a,b,c]	1	2	3[d]	4[e]	5[d]	6	7[f]	8	9	10	11	12
Just in time (JIT) principles						TQM[b]						
Quality management (QM)			Infrastructure[c]		JIT					TQM[a]		
Workforce management		Infrastructure[c]	Infrastructure[c]									
Setup time reduction	JIT system[b]	JIT[c]	JIT[c]	TBC[b]	JIT			JIT[b,c]	JIT[b]	JIT[a]	Lean[a]	X
Small lot size (reduction)	Flow[b]	JIT[c]								JIT[a]		X
Pull system (support)	Flow[b]											
Kanban/pull production	Flow[b]	JIT[c]	JIT[c]	TBC[b]	JIT			JIT[b,c]	JIT[b]	JIT[a]	Lean[a]	X
Equipment layout	Flow[b]		JIT[c]					JIT[b,c]	JIT[b]			X
(Continuous) flow										JIT[a]		X
Daily schedule adherence	JIT system[b]	JIT[c]	JIT[c]			TPM[b]		JIT[b,c]	JIT[b]			
Cellular manufacturing				TBC[b]	JIT					JIT[a]		X
Continuous improvement				TBC[b]						TQM[a]	Lean[a]	X
Statistical process control		TQM[c]	Quality management[c]						TQM[b]	TQM[a]		X
Group problem solving	JIT system[b]		Workforce management[c]									X
Training	JIT system[b]					TQM[b]	TPM[b]	Common[b,c]	HRM[b]	HRM[a]		X
Cross functional teams						TQM[b]	TPM[b]		HRM[b]	HRM[a]		X
Employee involvement				TBC[b,1]				Common[b,c]				X
Workforce commitment												
Preventive maintenance	JIT system[b]		JIT[c]	TBC[b]	JIT					TPM[a]		X
Product design (simplicity)	Flow[b]	TQM[c]	Infrastructure[c]					TQM[b,c]				
JIT delivery by suppliers	Supplier management[b]		JIT[c]					JIT[b,c]	JIT[b]			X
Supplier (quality) level	Supplier management[b]	Infrastructure[c]	Quality management[c]					TQM[b,c]				
Supplier relationship/involvement		Infrastructure[c]		TBC[b]		TQM[b]			TQM[b]		Lean[a]	X
Customer focus/involvement		TQM[c]				TQM[b]		TQM[b,c]	TQM[b]			X
JIT links with customers									JIT[b]			X

(1) Sakakibara et al. (1993); (2) Flynn et al. (1995a,b); (3) Sakakibara et al. (1997); (4) Koufteros et al. (1998); (5) Koufteros and Vonderembse (1998); (6) Dow et al. (1999); (7) McKone and Weiss (1999); (8) Cua et al. (2001); (9) Ahmad et al. (2003); (10) Shah and Ward (2003); (11) Li et al. (2005); (12) Current study.

[a] Used as an item to measure a first order construct.

[b] Used as first order construct to measure a second order construct.

[c] Reduced the first order construct to a single score.

[d] Measurement items are not included in the study.

[e] TBC: time based competition.

[f] TPM: total preventive maintenance; 1-not included by Nahm et al. (2003) in their measures of TBC.

measures, we refer to the survey questions used to represent individual lean practices/tools as manifest variables or items. When data reduction techniques were performed to collapse multiple items into a smaller number, we refer to each of them as a latent variable or a factor. Latent variables/factors may be of the first-order (i.e. when manifest variables are used for data reduction) or of second or higher order (i.e. when first order-factors are used for additional data reduction) and represent the underlying unobservable components of a lean system.

Although it was difficult to track items or factors from study to study, our literature review underscores three critical problems which have serious implications for theory building. First problem arises because some concepts have undergone a change in status over time. Preventive maintenance is a case in point. In most of the early research, it was used as one of the underlying dimensions of JIT (Sakakibara et al., 1993) but it is now established as an independent construct (McKone and Weiss, 1999) and is used to predict manufacturing performance (Cua et al., 2001; Shah and Ward, 2003). Second problem occurs when identical items are used to operationalize vastly different concepts and finally, the reverse case in which different items are used to operationalize the same construct. To illustrate the two issues, we look to the study by Koufteros et al. (1998) as an example. The authors conceptualize pull production as time-based manufacturing (TBM) and measure it using shop-floor employee involvement in problem solving, reengineering setups, cellular manufacturing, quality improvement efforts, preventive maintenance, and dependable suppliers. Their description equates pull production with TBM though the measurement items and factors (constructs) they use are similar to items that were used to represent JIT in previous work. Several researchers have used pull production and TBM interchangeably in subsequent research studies (Koufteros, 1999; Nahm et al., 2003). Whether pull production is really the same as TBM is a normative issue (and not central to the objective of our paper), but equating them results in a proliferation of additional indistinct terms that tend to obfuscate the substantive meaning of the constructs in question.

We found only two studies specifically related to measuring lean production (Shah and Ward, 2003; Li et al., 2005). Shah and Ward (2003) developed measures for lean *manufacturing* and operationalized it as bundles of practices related to total quality management, total preventive maintenance, and human resource management. They limit their analysis to four bundles that are oriented internally to reflect a firm's approach to managing its *manufacturing* operation. In contrast, Li et al. (2005) measure lean production very restrictively with five items that include set up time, small lot size, and pull production.

Overall, our literature review accentuates the expansive nature of conceptual definitions of lean production and the difficulty in discriminating its underlying components from each other and from the system. This indicates that both the conceptual and the operational space surrounding lean production are under-developed (Fig. 1). We conceptualize lean production more holistically by capturing both internal and external practices to better align lean production with its origins and develop an appropriate set of measures.

2.2. Lean production—a conceptual definition

Theory building requires that concepts are well-defined. However, Osigweh (1989) argues that it is not imperative that every concept in a theory is precisely defined, rather he suggests that the concepts that are defined are well-conceptualized and their definitions are

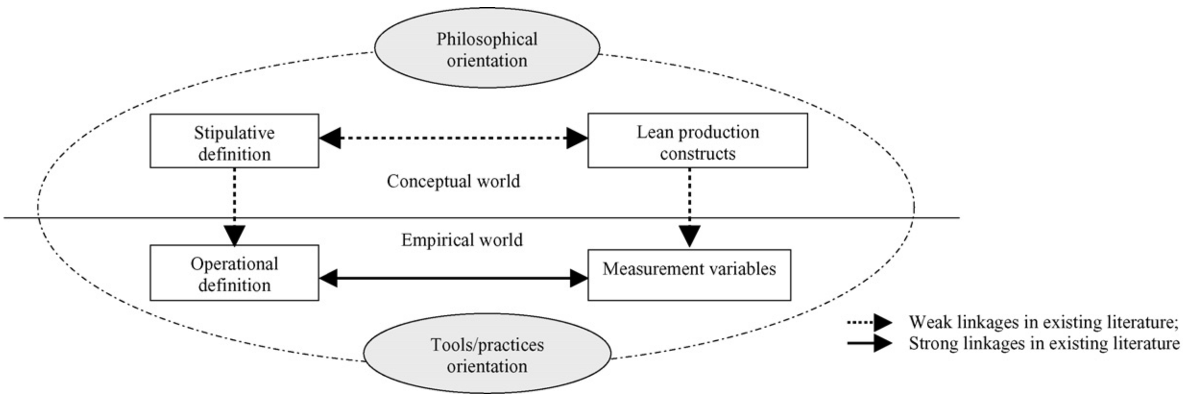

Fig. 1. Existing state of knowledge of the conceptual and empirical world as related to lean production (adapted from Priem and Butler, 2001).

sufficiently precise. Lean production is most frequently associated with elimination of waste commonly held by firms as excess inventory or excess capacity (machine and human capacity) to ameliorate the effects of variability in supply, processing time, or demand. According to Little's law (Anupindi et al., 1999), inventory in a system can be reduced by either maintaining excess capacity or lowering throughput time. Because excess capacity is a type of waste and is counter to lean production principles, lowering throughput time reliably to reduce inventory is preferred. This can be accomplished through continuous flow without frequent stop-and-go operations that are characteristic of batch and queue systems. Achieving this requires a flexible, dedicated and engaged work force. Therefore, to pursue lean production and minimize inventory, firms have to manage variability in supply, processing time, and demand (Hopp and Spearman, 2004; De Treville and Antonakis, 2006), which in turn require firms to effectively manage their social and technical systems simultaneously. We propose the following definition to capture the many facets of lean production.

> Lean production is an integrated socio-technical system whose main objective is to eliminate waste by concurrently reducing or minimizing supplier, customer, and internal variability.

Wacker (2004) suggests that a conceptual definition should show evidence of clarity, communicability, consistency, parsimony, differentiability, inclusivity, and exclusivity. We believe that our definition meets these criteria. We do not wish to propose one definition that fits all lean producers because such a definition will necessarily be narrow and context specific (Lorsch, 1980). Conversely, we do not want our definition to be so broad that it is a generality, encompassing nearly every organizational phenomenon (Putnam, 1978). Our definition of lean production will also help to bridge the gap between the differing philosophical and practice/tools perspectives witnessed in existing literature.

To manage variability in supply, processing time, and demand, firms pursuing lean production must become attentive to them and the underlying causes. Variability in supply occurs when suppliers fail to deliver the right quantity or right quality at the right time or the right place (Womack et al., 1990). This variability can be managed by creating a dependable and involved supplier base that consists of a few key suppliers with long term contracts. Other practices used to limit supplier variability include providing regular feedback on quality and delivery performance and providing training and development for further improvement.

Similarly, there are many practices and tools that minimize process time variability. For example, specifying work to its smallest detail enables line balancing and, thus, more predictable production quantities. A stringent quality assurance regimen reduces rework and results in less variability in process time. Cross-trained employees are able to step in for absent employees without disrupting flow, quality, and quantity of work (Monden, 1983, p. 3). Lean production includes these and many other practices and tools to minimize process time variability.

Finally, the effect of demand variability can ripple through the entire production process and cause havoc to daily production schedule. To counter the effects of demand variability, lean production focuses on *takt-time,* a measure of the amount of production required to meet customer demand, and production smoothing techniques such as ''heijunka'' to adapt to the changing demand (Monden, 1983, p. 2). Demand management may also be used to smooth fluctuations in patterns of demand over time intervals. Thus, the overarching philosophical imperative for waste reduction is accomplished through a variety of mutually reinforcing practices/tools which serve to reduce waste in very specific ways.

2.3. Lean production using the configurational lens

Configurations represent non-linear synergistic effects and higher-order interactions that cannot be represented with traditional bivariate or contingency relationships (Doty and Glick, 1994). Meyer et al. (1993) have defined configurations as ''any multi-dimensional constellation of conceptually distinct characteristics that commonly occur together''. Similarly, Miller and Friesen (1978) have noted that configurations are commonly occurring clusters of attributes or relationships that are internally cohesive. The common theme underlying configurations is the notion of *distinct characteristics that occur together.*

Lean production may be viewed as a configuration of practices/tools because the relationships among the elements of lean production are neither explicit nor precise in terms of linearity or causality. A configuration approach helps to explain how a lean system is designed from the interaction of its constituent elements taken as a whole, as opposed to designing the system one element at a time. From a theoretical standpoint, lean production is seen as a tightly coupled system where the constituent elements hold together in mutual dependence. It is the self-reinforcing effects of this kind of mutual dependence that contribute to the superior performance associated with lean production (Shah and Ward, 2003) on the one hand and make it rare, valuable and difficult to imitate by competitors on the other hand. Viewing lean production with a configurational lens provides us the logic that glues its multiple facets together.

3. Methods

The empirical objective of this study is to identify the dimensional structure underlying lean production and to develop reliable and valid scales to represent it. We adopted a comprehensive, multi-step approach (Fig. 2) during the development and validation process, followed by several studies in operations management (Koufteros et al., 1998; Nahm et al., 2003). Each of the steps is described briefly below.

3.1. Instrument development

Because our review of the related literature exposed considerable overlap in theoretical and operational concepts, we used past research to obtain general insights into measurement. We identified an initial list of items from the past literature to represent those aspects of lean production that eliminate or minimize different aspects of variability. The items were vetted through two steps to assure high research design quality. These included a structured interview with 10 practitioners to assess face and content validity of the scale items followed by a pre-test of the scale items with academics and practitioners. We provide an illustrative example of how the items composing a scale evolved during scale development in Appendix A. The final instrument (Appendix B) consisted of 48 items to reflect a comprehensive set of lean practices.

3.2. Sample selection

To avoid overly homogenous or heterogeneous samples, we considered the population of interest to be all manufacturing firms (SIC 20–39) who are identified as implementing lean production. We used two criteria to select firms in our sampling frame: (1) the firm had to belong to a manufacturing SIC code; (2) the firm's minimum number of employees had to exceed 100, as past research has shown that larger firms are more likely to implement lean production (Shah and Ward, 2003). A contact list from Productivity Inc., a firm involved with the consulting, training, and implementation of lean production systems, was selected as the primary source for developing a respondent profile. Productivity Inc. maintains a large database of high- and mid-level manufacturing executives from a diverse set of manufacturing companies. The resulting list is well-suited to the purpose of this study because it consists of a set of manufacturing firms that are at various stages of implementation of lean practices. The list was refined to eliminate duplicate records and incomplete job titles, resulting in a file with 750 records for pilot testing and 2616 records for large-scale validation.

3.3. Pilot study

Dillman's (2000) Total Design Method, commonly employed in operations management research (Koufteros et al., 1998; Nahm et al., 2003), was used with slight modifications to administer the survey. We made an initial contact with the respondent and followed up with a packet containing a cover letter and the survey for the pilot study. Instead of the two to three reminders advocated by Dillman (2000), we chose not to send any reminders for the pilot study. Sixty-three responses, corresponding to a response rate of 9% were used to assess initial reliability and to conduct exploratory data analysis. Although the pilot sample in our study appears small, it is, nonetheless, larger than samples used in other studies that have conducted a pilot study as part of their research design (Koufteros et al., 1998; Nahm et al., 2003). Additionally, executives contacted for the pilot study and the data collected from them were not used for the large-scale study, as the objective during the pilot study was only to identify a dimensional structure corresponding to lean production concept.

3.3.1. Data analysis

Exploratory data analysis was conducted as follows. First, descriptive statistics and missing item analysis were conducted for each of the 48 items. The results did not indicate any problems with the missing item analysis. Second, a Corrected Item to Total Correlation (CITC) score was calculated for each item to assess item reliability. Six items with CITC values below 0.30 were removed, and an additional reliability analysis was conducted. Five of the six excluded items were reverse-coded. Previous research has also indicated lower item reliability with reverse-coded items (Flynn et al., 1990).

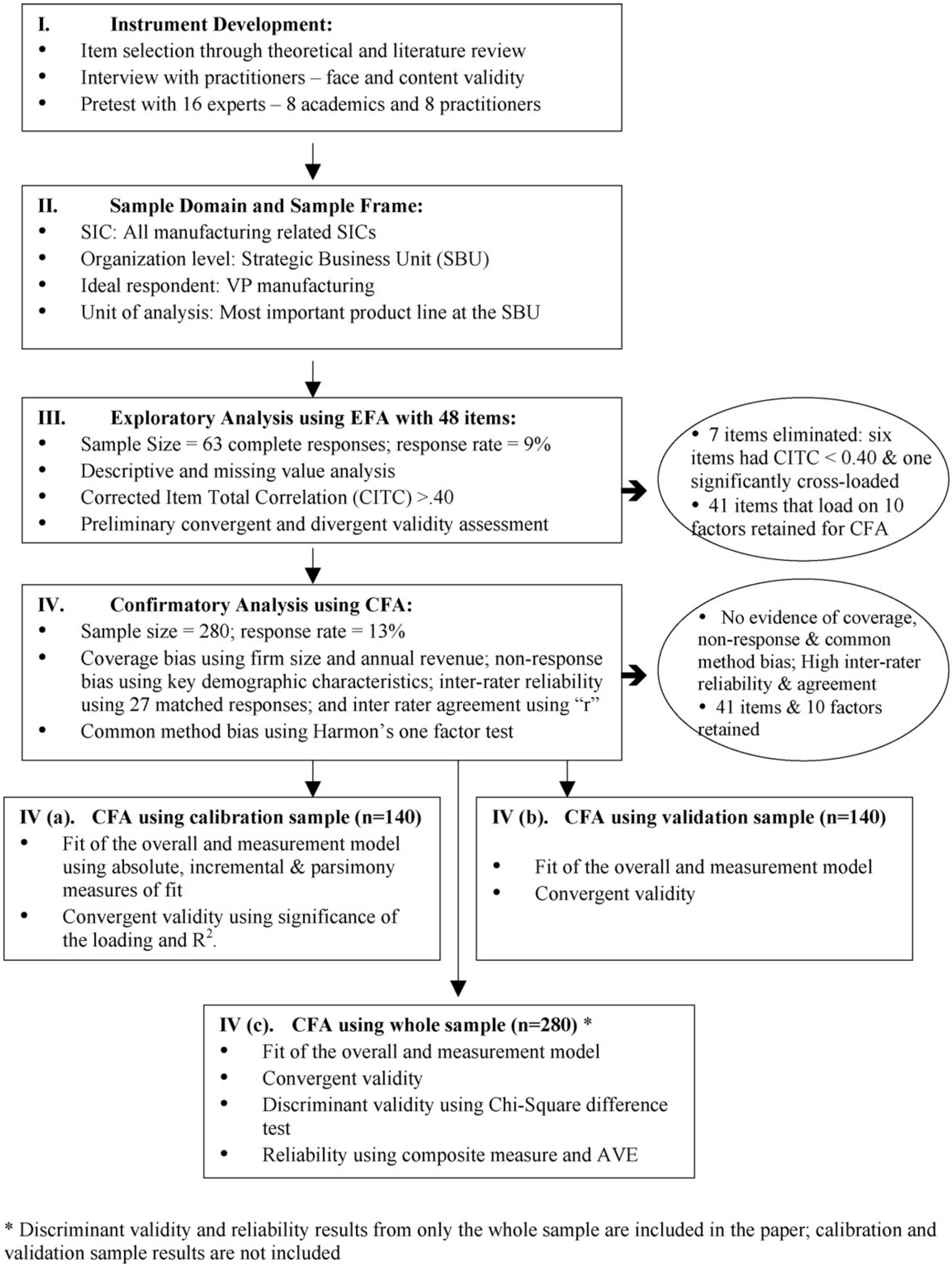

Fig. 2. Schematic representation of steps followed during scale development and validation.

Third, to assess divergent validity, the 42 items that were retained were subjected to an exploratory factor analysis (EFA). CF-VARIMAX, an oblique rotation, was used to extract common factors; maximum likelihood (ML) method was used to estimate the common factor model; and CEFA (comprehensive exploratory factor analysis, v.1) was used to conduct the analysis (available at http://quantrm2.psy.ohio-state.edu/browne/software.htm). A multiple-step iterative method was used to assess the appropriate number of common factors—2–12 factors were specified. Eigen values, communalities greater than 1, RMSEA, ECVI, and 90% confidence interval (CI) associated with RMSEA and ECVI were compared to decide the appropriate number of factors (for details on the iterative method, see Browne and Cudeck, 1992). Results[1] from iterative analysis suggested that a 10-factor solution was best. The factor structure and pattern of loadings obtained have face validity compared to past research. One item (Flow_05) significantly cross-loaded on three factors, therefore it was eliminated. Two pairs of items had large residual correlations with each other (Flow_01 with Flow_02 and SPC_01 with SPC_02), indicating that a correlation between error terms during confirmatory phase might be justified between them.

Finally, convergent validity was examined by conducting factor analysis at the individual factor level, and the internal consistency of each factor was examined using Cronbach Alpha. Results from factor level EFA indicated that all items had significant loadings on their respective factor, Eigen values exceeded 2, and the percent of variance explained ranged from 53 to 79%. Cronbach Alpha for each of the factors ranged between 0.73 and 0.86, indicating internal consistency. To summarize results from the exploratory phase, seven items were dropped during the iterative analysis. Based on the EFA results, 41 items are retained for the confirmatory phase and are hypothesized to load on 10 factors representing lean production.

4. Confirmatory phase: large-scale study

Satisfied by the initial reliability and validity of our measurement scales, we moved into the confirmatory phase of testing our survey instrument. Using Dillman's method, we made an initial contact with the respondents using a survey that was followed with two reminders, each sent a week apart. We received 295 responses for an effective response rate of 13.5%. The response rate exceeds recent survey-based research in operations management (7.47%, Nahm et al., 2003) and supply chain management (6.3%, Li et al., 2005). Fifteen responses were missing a substantial amount of data on the items used in this study therefore 280 responses were used for the analysis. Prior to conducting any data analysis, descriptive statistics were calculated for the 41 items. A pairwise *t*-test to compare the means for each of the items from the exploratory and confirmatory phase was conducted. No significant difference was found for any of the items.

4.1. Evaluating bias

Coverage bias was assessed by comparing the profile of the responding firms to the population of U.S. firms for number of employees and annual revenue and by computing χ^2-test statistic. Results indicated that our sample may be biased towards larger firms ($\chi^2 = 12.91$, d.f. = 3, p-value <0.005 for number of employees and $\chi^2 = 16.21$, d.f. = 3, p-value <0.001 for annual revenue). However, large firms implement lean practices significantly more often than do small firms (Shah and Ward, 2003); therefore bias in favor of large firms should not impact the results because selecting a sample to maximize the variance of measured variables relevant to the constructs of interest is highly recommended (Fabrigar et al., 1999). Non-response bias was assessed by comparing respondents to non-respondents on key demographic characteristics (Filion, 1976). These groups were compared on the basis of annual revenue, number of employees, and sales volume with the data provided by Productivity Inc. No significant differences (lowest p-value >0.20) were found between the two groups, thus ameliorating concern about possible non-respondent bias.

Second raters were solicited from a sub-sample ($n = 108$) of responding companies. Twenty-seven complete responses were received from second raters, corresponding to a response rate of almost 25% (27/108). Following Boyer and Verma (2000), we measured both inter-rater reliability and inter-rater agreement. Whereas inter-rater reliability is an index of consistency and is generally correlational in nature, inter-rater agreement refers to the interchangeability among raters and addresses the extent to which raters make essentially the same ratings (Kozlowski and Hattrup, 1992). To estimate inter-rater reliability, Pearson's bivariate correlation and inter-class correlation (ICC) were computed on 41 items from the 27 matched responses. All of the

[1] Available from the authors upon request.

items had positive and significant correlation, and the corresponding p values were all below 0.05. ICC was estimated by the method suggested by Shrout and Fleis (1979). The ICC exceeded 0.60 and the F-value was greater than 3.51 ($p < 0.000$) for each of the items. Both the tests indicate a high level of consistency between the two raters. Inter-rater agreement (r), a measure of the proportion of the observed variance to expected variance, was calculated using the method described by Finn (1970). The r values ranged from 0.65 to 0.75. Although no absolute standard for this measure has been established, Boyer and Verma (2000) suggest that the values achieved may be acceptable. Therefore, we conclude that both inter-rater reliability and inter-rater agreement are acceptable for each of the items.

Common method bias is present when correlations between measures can be explained by the fact that the same individual provides the responses for all measurement scales rather than by any true relationship between the constructs. Because common method bias presents a greater cause for concern in self-report studies, Harmon's one-factor test was conducted. Common method bias is present if a factor analysis using all relevant measurement items results in a single factor (Podaskoff and Organ, 1986; Miceli et al., 1991). We performed an exploratory factor analysis with no rotation and found 10 factors with Eigen values greater than one. While this does not provide conclusive evidence of absence of common method bias, it suggests that any common method bias that does exist is unlikely to be problematic. Consequently, a reliability analysis was conducted on the 41 items to assess corrected-item total correlation (CITC) scores and Cronbach Alpha. Each of the items had a CITC score above 0.40, and the Cronbach Alpha was 0.93. Thus, the set of 41 items was carried forward to the CFA.

4.2. *Confirmatory factor analysis*

A CFA involves the estimation of an *a priori* measurement model, where the observed variables are mapped onto the latent constructs according to theory and prior testing by the researcher. The final model extracted during EFA was used to specify the relationships. We used a covariance matrix of the 41 items to input data, maximum likelihood method to estimate the model, and LISREL 8.51 to conduct the analysis. Because each of the latent variables was measured using three or more manifest variables, our model is identified (Long, 1983). In addition to *a priori* specification of the factor structure, CFA provides a more rigorous test of convergent and discriminant validity than the more traditional multitrait-multimethod analysis (Campbell and Fiske, 1959). Our approach to examining validity, unidimensionality, and reliability is described below.

First, convergent validity and an item-level reliability test is conducted to assess how a particular item behaves within the block of items intended to measure a latent variable. Anderson and Gerbing (1988) suggest that evidence of convergent validity exists if the manifest variable loads significantly (t-value >2.58, $p < 0.01$) on its respective latent variable. The proportion of variance explained (R^2) in the manifest variable that is accounted for by the latent variable influencing it can be used to estimate the reliability of a particular item.

Second, we assessed the model fit and unidimensionality of the model. Model fit was evaluated using multiple absolute, incremental, and parsimonious measures of fit which provide different aspects to answering the question "How well do the relationships estimated by the model match the observed data?" Absolute measures of fit assess how well an *a priori* model reproduces the sample data; incremental fit measures assess the incremental fit of the model compared to a null or worst-case model; and parsimonious fit measures assess the parsimony of the proposed model by evaluating the fit of the model versus the number of estimated coefficients needed to achieve the level of fit (Hair et al., 1998). Because many fit indices are affected by sample size (e.g. GFI, NFI and AGFI) and others by the ratio of manifest variables per latent variable (e.g. NNFI and CFI), Shah and Goldstein (2006) suggest reporting a broad set of fit indices. We include multiple measures of fit to reach meaningful conclusions with regards to model fit and report our results and the recommended cutoffs for each of the measures. We also examine the magnitude of standardized residuals and modification indices to estimate any mis-specification in the model.

Finally, discriminant validity is assessed by constructing models for all possible pairs of latent variables: first, a model is run where the covariance between any two latent variables is fixed to one; second, another model is run where the covariance between them is free to assume any value; and third, the significance of the χ^2 difference between the two models is computed. Constraining the covariance between two latent variables is similar to stating that they are unidimensional and not unique or distinct. Because the free model is nested within the constrained model, the χ^2 difference can be tested for significance with one degree of freedom. A significant χ^2 difference indicates that constraining pairwise covariance to one

Table 4
Measurement model fit for the calibration and validation samples

Item name		LV	Calibration sample (n=140)		Validation sample (n=140)		Whole sample (n=280)	
			λ_i (S.E.)[a]	R^2	λ_i (S.E.)[a]	R^2	λ_i (S.E.)[a]	R^2
Suppfeed_01	<-	SUPPFEED	0.69 (0.079)	0.49	0.65 (0.086)	0.42	0.67 (0.058)	0.46
Suppfeed_04	<-	SUPPFEED	0.78 (0.076)	0.62	0.59 (0.087)	0.36	0.70 (0.057)	0.49
Suppfeed_05	<-	SUPPFEED	0.87 (0.077)	0.71	0.77 (0.080)	0.66	0.82 (0.055)	0.69
SuppJIT_01	<-	SUPPJIT	0.55 (0.089)	0.39	0.54 (0.087)	0.32	0.55 (0.062)	0.32
SuppJIT_02	<-	SUPPJIT	0.77 (0.084)	0.59	0.65 (0.086)	0.45	0.70 (0.060)	0.50
SuppJIT_03	<-	SUPPJIT	0.57 (0.084)	0.36	0.62 (.0090)	0.38	0.60 (0.061)	0.38
Suppdevt_01	<-	SUPPDEVT	0.56 (0.082)	0.33	0.52 (0.087)	0.27	0.55 (0.060)	0.31
Suppdevt_02	<-	SUPPDEVT	0.59 (0.082)	0.36	0.70 (0.081)	0.49	0.65 (0.058)	0.49
Suppdevt_03	<-	SUPPDEVT	0.41 (0.090)	0.18	0.38 (0.086)	0.15	0.40 (0.062)	0.19
Suppdevt_04	<-	SUPPDEVT	0.64 (0.070)	0.45	0.69 (0.085)	0.44	0.65 (0.058)	0.44
Suppdevt_05	<-	SUPPDEVT	0.48 (0.082)	0.29	0.57 (0.089)	0.30	0.52 (0.060)	0.30
Suppdevt_06	<-	SUPPDEVT	0.48 (0.082)	0.33	0.62 (0.086)	0.36	0.55 (0.060)	0.36
Custinv_01	<-	CUSTINV	0.40 (0.077)	0.24	0.56 (0.083)	0.30	0.50 (0.057)	0.26
Custinv_03	<-	CUSTINV	0.52 (0.076)	0.34	0.65 (0.078)	0.42	0.59 (0.055)	0.36
Custinv_04	<-	CUSTINV	0.87 (0.060)	0.84	0.93 (0.071)	0.80	0.90 (0.046)	0.84
Custinv_05	<-	CUSTINV	0.86 (0.062)	0.84	0.94 (0.069)	0.84	0.90 (0.046)	0.85
Custinv_06	<-	CUSTINV	0.63 (0.070)	0.49	0.78 (0.079)	0.54	0.70 (0.053)	0.50
Pull_01	<-	PULL	0.90 (0.074)	0.72	0.63 (0.069)	0.48	0.77 (0.051)	0.61
Pull_02	<-	PULL	0.91 (0.069)	0.79	0.79 (0.067)	0.69	0.86 (0.048)	0.75
Pull_03	<-	PULL	0.97 (0.065)	0.91	0.89 (0.063)	0.89	0.93 (0.045)	0.89
Pull_04	<-	PULL	0.64 (0.079)	0.41	0.68 (0.073)	0.51	0.66 (0.053)	0.44
Flow_01	<-	FLOW	0.48 (0.079)	0.28	0.50 (0.079)	0.22	0.47 (0.061)	0.23
Flow_02	<-	FLOW	0.50 (0.083)	0.31	0.61 (0.085)	0.35	0.54 (0.059)	0.35
Flow_03	<-	FLOW	0.78 (0.079)	0.66	0.95 (0.075)	0.87	0.87 (0.055)	0.78
Flow_04	<-	FLOW	0.68 (0.078)	0.53	0.74 (0.082)	0.51	0.70 (0.057)	0.51
Setup_01	<-	SETUP	0.87 (0.084)	0.73	0.79 (0.072)	0.68	0.82 (0.055)	0.69
Setup_02	<-	SETUP	0.55 (0.082)	0.33	0.79 (0.080)	0.59	0.66 (0.058)	0.45
Setup_03	<-	SETUP	0.50 (0.083)	0.27	0.72 (0.081)	0.50	0.61 (0.058)	0.38
SPC_01	<-	SPC	0.60 (0.079)	0.39	0.58 (0.083)	0.34	0.60 (0.058)	0.38
SPC_02	<-	SPC	0.64 (0.079)	0.43	0.78 (0.076)	0.61	0.72 (0.055)	0.54
SPC_03	<-	SPC	0.74 (0.074)	0.59	0.80 (0.077)	0.62	0.77 (0.053)	0.61
SPC_04	<-	SPC	0.63 (0.074)	0.42	0.77 (0.080)	0.56	0.70 (0.055)	0.56
SPC_05	<-	SPC	0.66 (0.071)	0.53	0.69 (0.086)	0.41	0.67 (0.056)	0.45
Empinv_01	<-	EMPINV	0.80 (0.074)	0.63	0.73 (0.074)	0.57	0.77 (0.052)	0.61
Empinv_02	<-	EMPINV	0.70 (0.069)	0.58	0.79 (0.079)	0.58	0.75 (0.053)	0.58
Empinv_03	<-	EMPINV	0.81 (0.071)	0.68	0.83 (0.073)	0.69	0.82 (0.051)	0.69
Empinv_04	<-	EMPINV	0.72 (0.075)	0.54	0.69 (0.077)	0.49	0.71 (0.054)	0.52
TPM_01	<-	TPM	0.59 (0.087)	0.43	0.69 (0.083)	0.49	0.65 (0.061)	0.49
TPM_02	<-	TPM	0.68 (0.080)	0.50	0.75 (0.087)	0.53	0.73 (0.059)	0.53
TPM_03	<-	TPM	0.88 (0.083)	0.72	0.65 (0.082)	0.45	0.74 (0.059)	0.54
TPM_04	<-	TPM	0.55 (0.085)	0.32	0.63 (0.088)	0.38	0.57 (0.062)	0.38

[a]Standardized factor loading (standard error).

will result in a significant model misfit, and, therefore, two unique latent variables are necessary to explain the factor structure. Reliability estimation is conducted last because, in the absence of a valid construct, reliability is almost irrelevant (Koufteros, 1999). Cronbach Alpha, composite reliability, and average variance extracted (AVE) are used to assess internal consistency.

We took extra precautions to develop a reliable and valid set of measures for lean production. In order to test the measurement model, we randomly divided our total sample ($n = 280$) into two equal halves ($n = 140$)—a calibration sample and a validation sample. The sample size for the two samples exceeds the minimum sample size required to obtain a 0.80 statistical power with 732 degrees of freedom at alpha of 0.05 (MacCallum et al., 1996). The calibration sample was used in the initial model testing, while the validation sample was used to confirm the model. We estimated all measures of reliability, validity, and unidimensionality for calibration, validation, and whole sample separately. We report convergent validity, item reliability, model fit, and estimates of unidimensionality for calibration, validation, and the whole sample. In contrast, discriminant validity and reliability are reported for the whole sample only (although they were conducted for the calibration and validation samples also) to avoid redundancy.

4.2.1. CFA results using calibration sample

In the first iteration, the factor loadings for each of the 41 items were significantly larger than their standard errors, and the associated *t*-values exceeded 3.29 ($p < 0.001$). All the fit statistics were within the accepted range, but an analysis of the standardized residuals and modification indices suggested that two pairs of manifest variables had excessively large standardized residuals (>7) and modification indices (>30) associated with them. These are the same manifest variables (Flow_01 with Flow_02 and SPC_01 with SPC_02) that had large residuals during EFA. Therefore, another model was specified by including a correlation between the error terms for these variables. An alternative would be to drop two of the four items from the model, but we decided to keep all four items in order to represent their respective domains more completely.

The results from the second iteration show that the factor loadings for all 41 of the items are significantly larger than their standard errors resulting in *t*-values that exceed 3.29 ($p < 0.001$). The variance explained range was between 0.18 and 0.91 (Table 3). The multiple measures of model fit indicate a mixed picture: RMSEA, 90% confidence interval associated with RMSEA, RMR, and normed χ^2 indicate a good to excellent fit, but NNFI and CFI are at or below the recommended level (Table 5; columns 4 and 5). The proportion of absolute standardized residuals > $|2.58|$ is 6.35% (54 out of 851). A value of $|2.58|$ corresponds to the area beyond the ± 2 standard deviations from the average standardized residual or the values lying in the extreme 5% of the distribution. All modification indices are below 20.

4.2.2. CFA results using validation sample

The measurement model incorporating the modifications described in Section 4.2.1 was retested using the validation sample. Results for the convergent validity from the validation sample are reported in Table 4 (columns 6 and 7) and measures of model fit in Table 5

Table 5
Absolute, Incremental and Parsimonious measures of fit for the calibration and validation samples

Measures of fit	Statistic measure	Calibration sample ($n = 140$)	Validation sample ($n = 140$)	Whole sample ($n = 280$)	Recommended value for close or acceptable fit
Absolute	χ^2-Test statistic (d.f.)	1114.59 (732)	1052.78 (732)	1178.01 (732)	NA
	Root mean square error of approximation (RMSEA), point estimate	0.055	0.048	0.047	≤0.08
	RMSEA, 90% confidence interval	(0.047; 0.063)	(0.039; 0.056)	(0.042; 0.052)	(0.00; 0.08)
	p value H_0: close fit (RMSEA ≤ 0.05)	0.13	0.66	0.86	≥0.05
	Standardized root mean square residual (RMR)	0.078	0.073	0.062	≤.10
Incremental	Non-normed fit index (NNFI)	0.85	0.86	0.91	≥0.90
	Comparative fit index (CFI)	0.86	0.88	0.92	≥0.90
	Incremental fit index (IFI)	0.87	0.88	0.91	≥0.90
Parsimonious	Normed χ^2 (χ^2/d.f.)	1.52	1.44	1.61	≤3.0
	Parsimony normed fit index (PNFI)	0.62	0.62	0.72	≥0.70

(column 4). The pattern and size of loadings and the variance explained is similar to those in the calibration sample. RMSEA, 90% confidence interval associated with RMSEA, RMR, and Normed χ^2 indicate a good to excellent fit, but NNFI and CFI continue to be at or below the recommended level. The proportion of absolute standardized residuals $> |2.58|$ is 4.35% (37 out of 851) and all modification indices are less than 10. These results indicate that the validation sample explains the relationships in the final measurement model well. These results also indicate that the two samples exhibit *invariance of form* (i.e. using the same mapping of manifest variables to latent variables in two sub-samples is appropriate).

4.2.3. CFA results using the whole sample

The path loadings between the item and its corresponding factor were all positive and significant at $p < 0.001$, and the value of the path loading ranged between 0.50 and 0.94 (Table 3, columns 7 and 8). Amount of variance explained by each of the items ranged from 0.18 to 0.90; the average amount of variance explained equals 0.50. In terms of model fit, all the measures improved greatly with the use of the whole sample due to increased sample size. The RMSEA value equaled 0.049, the p-value associated with the null hypothesis of close fit (0.860) indicates that it could not be rejected, and the data explains 86% of the variance in the hypothesized model. All fit measures were at or above the recommended value. The proportion of absolute standardized residuals $> |2.58|$ is 9.28% (79 out of 851), and all modification indices are <10.

To assess discriminant validity, we made 45 pairwise comparisons between the fixed and the free models. The smallest χ^2 difference was 8.38 ($p < 0.003$), which is significant with one degree of freedom; all other comparisons were significant at lower p-values (Table 6). Cronbach Alpha exceeded 0.70 for all 10 factors and composite reliability exceeded 0.70 in nine out of 10 factors. AVE exceeds the 0.50 threshold value for 6 out of 10 factors. As a set, these results indicate that construct reliability depends on the measure employed to assess it: Cronbach Alpha designates adequate reliability for all the constructs, composite reliability and variance extracted also present an acceptable picture.

Table 6
Correlations, reliability and discriminant validity for the whole sample ($n = 280$)*

	Latent Variable	# of items	1	2	3	4	5	6	7	8	9	10
1	SUPPFEED	3	0.77 (.53)	32.4	38.3	64.6	34.7	21.6	29.4	38.4	14.8	56.8
2	SUPPJIT	3	0.53	0.66 (.39)	48.7	69.1	14.6	27.2	33.5	41.0	30.8	52.6
3	SUPPDEVT	6	0.65	0.76	0.74 (.33)	61.8	41.8	43.5	29.8	31.8	29.8	57.5
4	CUSTINV	5	0.29	0.22	0.43	0.85 (.55)	53.6	62.3	44.6	46.3	66.3	66.6
5	PULL	4	0.26	0.56	0.38	0.22	0.88 (.66)	9.5	25.2	41.4	27.8	29.6
6	FLOW	4	0.36	0.43	0.35	0.15	0.49	0.75 (.45)	15.5	32.2	18.8	30.4
7	SETUP	3	0.38	0.46	0.52	0.36	0.48	0.55	0.75 (.51)	11.2	8.4	44.5
8	SPC	5	0.42	0.45	0.66	0.44	0.31	0.37	0.61	0.86 (.60)	16.1	49.1
9	EMPINV	4	0.48	0.49	0.59	0.28	0.42	0.42	0.64	0.62	0.83 (.50)	46.0
10	TPM	4	0.19	0.30	0.26	0.12	0.31	0.27	0.19	0.25	0.17	0.77 (.45)

* The lower triangle shows correlations; all correlations are significant at $p < 0.001$. Composite reliability and average variance extracted (in parentheses) are on the diagonal in closed boxes. The upper triangle shows the difference in χ^2-test statistic between a fixed and a free CFA model; all χ^2 differences are significant at $p < 0.003$.

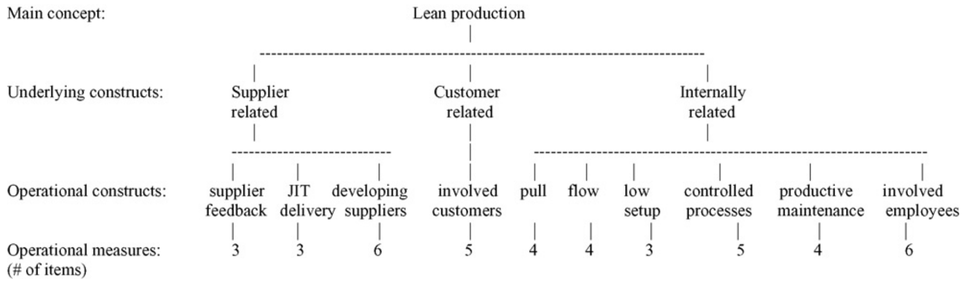

Fig. 3. Conceptual and empirical mapping as proposed and validated in the current study.

5. Discussion and conclusion

This research takes an initial step toward clarifying the concept of lean production and develops and validates a multi-dimensional measure of lean production. Following Whetten (1989), we organize the discussion of our results into three sections: *what* is lean production (i.e. identify critical factors), *how* are the various factors of lean production related to each other, and *why* are they related.

5.1. "What" is lean production?

In this research, we propose a conceptual definition of lean production and derive an operational measure from the content and objectives of its historical roots in TPS. To that end, we identified 48 practices/tools to represent the operational space surrounding lean production. Using a multi-step construct development method, we distill the measurement items into 10 factors which also map onto the conceptual definition well. Fig. 3 summarizes this mapping. Of the 10 factors identified in this study, three measure supplier involvement, one measures customer involvement, and the remaining six address issues internal to the firm. Together, these 10 factors constitute the operational complement to the philosophy of lean production and characterize 10 distinct dimensions of a lean system. They are:

1. SUPPFEED (supplier feedback): provide regular feedback to suppliers about their performance
2. SUPPJIT (JIT delivery by suppliers): ensures that suppliers deliver the right quantity at the right time in the right place.
3. SUPPDEVT (supplier development): develop suppliers so they can be more involved in the production process of the focal firm.
4. CUSTINV (customer involvement): focus on a firm's customers and their needs.
5. PULL (pull): facilitate JIT production including kanban cards which serves as a signal to start or stop production.
6. FLOW (continuous flow): establish mechanisms that enable and ease the continuous flow of products.
7. SETUP (set up time reduction): reduce process downtime between product changeovers.
8. TPM (total productive/preventive maintenance): address equipment downtime through total productive maintenance and thus achieve a high level of equipment availability.
9. SPC (statistical process control): ensure each process will supply defect free units to subsequent process.
10. EMPINV (employee involvement): employees' role in problem solving, and their cross functional character.

5.2. "How" are the factors of lean production related?

The 10 factors derived during empirical analysis are positively and significantly correlated with each other ($p < 0.001$), thereby providing support to the multi-dimensional and integrated nature of lean production systems (Table 6). As a set, the statistical and empirical results associated with the CFA model suggest that lean production can be represented with 10 factors where each factor represents a unique facet. The high inter-correlation between the factors lends further support to the "configuration" argument and suggests that managers are able to discern the close relationship and yet make distinctions between them. More specifically, our results indicate that practicing managers recognize the contribution of each individual factor and their collective importance when pursuing lean production. These results allow us to build on existing agreement related to "bundles of lean practices" (Shah and Ward, 2003) and, at the same time, to add clarity to the confusing array of terms and concepts associated with lean production.

The correlations between factors range from 0.77 (between SUPPJIT and SUPPDEVT) to 0.12 (between TPM and CUSTINV). A closer inspection of the correlation matrix revealed that TPM is the least associated with other scales. This pattern in the correlation matrix was also observed in Sakakibara et al. (1993), where preventive maintenance exhibited low and insignificant correlations with five other scales. However we could not examine this relationship in other published research because none of them included the correlation matrix between TPM and other factors (McKone and Weiss, 1999; Cua et al., 2001; Shah and Ward, 2003). Even so, the CFA results suggest that a well-developed lean strategy will include many different lean practices. Therefore, a state of the art implementation will require firms to exert considerable effort along several dimensions simultaneously.

5.3. "Why" are the factors of lean production related?

Lean production is an integrated system composed of highly inter-related elements. In explaining inter-relationships, researchers frequently rely on the statistical significance of the empirical results. However, statistical significance is a necessary but not a sufficient condition to explain the inter-relationships in a system. Researchers must also judge the reasonableness of the logic used to explain the inter-relatedness of the elements because reasonable logic provides the theoretical glue that holds a model together (Whetten, 1989). We explain our logic below.

The main objective of lean production is to eliminate waste by reducing or minimizing variability related to supply, processing time, and demand. Reducing variability related to only one source at a time helps a firm in eliminating only some of the waste from the system; not all waste can be addressed unless firms can attend to each type of variability concomitantly. That is, processing time variability cannot be eliminated unless supply and demand variability is also reduced. For instance, variability in setup times and delivery schedule by suppliers both contribute to firms holding excess inventory in order to prevent starving downstream work stations. But reducing setup time variability alone does not eliminate excess inventory from the system, because firms will continue to hold excess inventory to accommodate variability in supplier delivery. To reduce excess inventory of all types, firms will have to secure reliable suppliers in addition to developing a reliable process.

The 10 underlying factors/dimensions of lean production proposed here jointly enable firms to address variability in the following manner. To facilitate continuous flow (FLOW), products are grouped according to product families, and equipment is laid out accordingly; and to prevent frequent stop-and-go operations, equipment undergoes frequent and regular preventive maintenance (TPM). Closely grouped machines and the similarity of products allow employees to identify problems while cross-trained, self-directed teams of workers are able to resolve problems more quickly and effectively (EMPINV). Actively involved customers (CUSTINV) enable firms to predict customer demand accurately. Reduced setup times (SETUP) and stricter quality assurance (SPC) allow firms to predict process output more exactly. To produce the kind of units needed, at the time needed, and in the quantities needed, firms use kanban and pull production systems (PULL), which require that suppliers deliver sufficient quantities of the right quality product at the right time. This JIT delivery by suppliers (SUPPJIT) is predicated on providing suppliers with regular feedback on quality and delivery performance (SUPPFEED), and providing training and development for further improvement (SUPPDEVT). Because no firm has infinite resources to expend, the supplier base needs to be limited to a few key suppliers with whom firms can have long term relationships rather than short term contracts.

It is the complementary and synergistic effects of the 10 distinct but highly inter-related elements that give lean production its unique character and its superior ability to achieve multiple performance goals. While each element by itself is associated with better performance, firms that are able to implement the complete set achieve distinctive performance outcomes that can result in sustainable competitive advantage. Sustainability of advantage follows from the difficulty in implementing several aspects of lean simultaneously. Because simultaneous implementation of so many elements is difficult to achieve, it is also difficult to imitate.

5.4. Contributions

We make three substantive contributions to existing research. First, viewing lean production in its historical context with an evolutionary lens helps to reconcile the overlap among its various components. We argue that, viewed separately, none of the components are equivalent to the system, but together they *constitute* the system. Lean production is not a singular concept, and it cannot be equated solely to waste elimination or continuous improvement, which constitute its guiding principles, nor to JIT, pull production, kanban, TQM, or employee involvement, which make-up some of its underlying components. Lean production is conceptually multifaceted, and its definition spans philosophical characteristics that are often difficult to measure directly. Further, the practices/tools used to measure lean production, even when associated uniquely with a single component, indicate mutual support for multiple components. By juxtaposing the historical evolution of lean production and the perspective used in describing it, we can begin to understand the multiplicity of terms associated with lean production and attempt to resolve some of the confusion surrounding it.

Second, we propose a conceptual definition of lean production which captures the integrated nature of lean systems. Our definition includes both the people and the process components on one hand, and internal (related to the firm) and external (related to supplier and customer) components on the other hand. In this sense, our definition of lean production highlights mechanisms needed to achieve the central objective of waste elimination. This definition maximizes the potential for concept-traveling so that lean production can precisely fit a variety of applications (Osigweh, 1989). Yet, it minimizes the problem of concept-stretching, or broadening the concept's meaning beyond reason. In order for a system to be lean, it has to address not only variability reduction, but also the specific operationalization of supplier and customer relationships which may differ depending on the unit of analysis.

Finally, we develop an operational measure of lean production and provide a framework that identifies its most salient dimensions. Our operational measure is more comprehensive than other measures observed in literature as it reflects the lean landscape more broadly by including both internal and external dimensions. An empirical test of our operational measures suggests that it is reliable and meets established criteria for assessing validity. We identified a broad set of items that can be distilled into fewer components to represent multiple facets of a lean production system. In identifying 10 dimensions of lean production, we help to establish its underlying dimensional structure. Specifically, we characterize lean production with 10 unique sub-dimensions, and in our attempt to resolve the confusing array of concepts and measurement schemes witnessed in the previous literature, we show that concepts and measurement scales change with time. This is consistent with Devlin et al.'s (1993) argument that there are no ''best'' or ''perfect'' scales and Schmalensee (2003), who argued that the choice of scales changes with research objective.

The empirically validated measurement instrument we provide here is useful for researchers who are interested in conducting survey research related to lean production systems. This instrument will allow the researchers to assess the state of lean implementation in firms and to test hypotheses about relationships between lean production and other firm characteristics that affect firm performance. The findings provide guidance for empirical research seeking parsimony in data collection. To adequately measure lean production, instrumentation should reflect all 10 underlying constructs. Additionally, the study provides a tool for managers to assess the state of lean production in their specific operations. For instance, scales developed here may be used by managers to self-evaluate their progress in implementing lean production. The findings also suggest that every one of the 10 dimensions of lean production is an important contributor and that none should be eliminated.

The framework forms a foundation for research in lean production and should prove helpful in enabling researchers to agree on a definition. It is imperative to come to agreement on both a conceptual definition and an operational measure because, if history is any guide, old concepts will continue to evolve and we in the academic community will lag farther behind practice. The new and emerging concept of ''lean-sigma'' is a case in point. Lean-sigma is being forwarded as a management philosophy based on integrating lean production principles and practices with Six Sigma tools. If we cannot consistently define lean production, how can we differentiate it from other management concepts and verify its effectiveness to practicing managers? It is our intention to contribute to scholarly agreement on such a stipulative definition and to the emerging academic literature related to lean production (Narasimhan et al., 2006; De Treville and Antonakis, 2006; Hopp and Spearman, 2004). We intend that our operational measure of lean production will complement the conceptual definition presented earlier.

6. Limitations and future research

Our findings may be limited by the specific research design that was used. First, we used single key informants to collect data. While multiple informants are typically recommended to validate the information obtained, it is difficult to get multiple informants to agree to participate. To help offset the single informant concern we used data from twenty-seven second raters to evaluate inter-rater reliability and inter-rater agreement. Despite a limited sample of second raters, neither metric indicated any significant difference on the items used in this research. Therefore, the single informant should have minimal (if any) impact on the validity of the results.

Another limitation of our study concerns our specific results. As such, research related to construct validation seeks to examine the degree of correspondence between the results obtained using a particular measurement scheme and to assign the meaning attributed to those results. Strong statistical results obtained from CFA provide compelling evidence to the factor structure. However, the factor structure needs to be reexamined in future research. If replicated, future researchers and managers can use them in studying and implementing lean production in isolation or in conjunction with other concepts such as Six Sigma. Successful replication would allow future researchers to eliminate measuring individual practices and considerably shorten the measurement instrument because the 10 factors will be sufficient to represent the underlying dimensions of lean production. This is consistent with Boyer and Pagell's (2000) suggestion of eliminating lower level variables when these are included in higher-level constructs.

Despite these limitations, our study contributes to the body of theory presented in the literature. In particular, we hope that the measurement scales developed here will provide a foundation for furthering lean production research that is more consistent. The empirical study of lean production systems is in its infancy, and moving it forward requires reliable and valid scales.

Appendix A. Illustrative example of construct development

Content domain: pull production	Practitioner interview	Pretest	Pilot test	Large scale
Production is "pulled" by the shipment of finished goods	X	X	X	X
Production at stations is "pulled" by the current demand of the next station	X	X	X	X
Vendors fill our kanban containers	X, A			
Our suppliers deliver to us in kanban containers	X, A			
We use kanban, squares, or containers of signals for production control	X	X	X	X
We use a kanban pull system for production control	X, B			
We use a "pull" production system	X, B	X	X	X
We use a pull system rather than work orders	X, B			

X: measurement items included in the study during the various stages; X, A: during structured interview, practitioners noted that these items (1) did not belong to pull production domain (2) belonged to JIT delivery by suppliers; (3) were duplicates of each other. They recommended using "our suppliers deliver to us on JIT basis" as alternate wording. X, B: practitioners recommended using one of the three items because the items duplicated the information.

Appendix B. Scales

Please indicate the extent of implementation of each of the following practices in your plant. (1) no implementation; (2) little implementation; (3) some implementation; (4) extensive implementation; (5) complete implementation.

Item no.	Item label	Final CITC score
Suppfeed_01	We frequently are in close contact with our suppliers	0.40
Suppfeed_02	Our suppliers seldom visit our plants (reverse coded) [a]	
Suppfeed_03	We seldom visit our supplier's plants (reverse coded) [a]	
Suppfeed_04	We give our suppliers feedback on quality and delivery performance	0.54
Suppfeed_05	We strive to establish long-term relationship with our suppliers	0.45
SuppJIT_01	Suppliers are directly involved in the new product development process	0.48
SuppJIT_02	Our key suppliers deliver to plant on JIT basis	0.48
SuppJIT_03	We have a formal supplier certification program	0.45
Suppdevt_01	Our suppliers are contractually committed to annual cost reductions	0.51
Suppdevt_02	Our key suppliers are located in close proximity to our plants	0.52
Suppdevt_03	We have corporate level communication on important issues with key suppliers	0.41
Suppdevt_04	We take active steps to reduce the number of suppliers in each category	0.54
Suppdevt_05	Our key suppliers manage our inventory	0.40
Suppdevt_06	We evaluate suppliers on the basis of total cost and not per unit price	0.47
Custinv_01	We frequently are in close contact with our customers	0.40
Custinv_02	Our customers seldom visit our plants (reverse coded) [a]	
Custinv_03	Our customers give us feedback on quality and delivery performance	0.48
Custinv_04	Our customers are actively involved in current and future product offerings	0.42
Custinv_05	Our customers are directly involved in current and future product offerings	0.43
Custinv_06	Our customers frequently share current and future demand information with marketing department	0.42
Custinv_07	We regularly conduct customer satisfaction surveys [a]	
Pull_01	Production is "pulled" by the shipment of finished goods	0.47
Pull_02	Production at stations is "pulled" by the current demand of the next station	0.50
Pull_03	We use a "pull" production system	0.54
Pull_04	We use Kanban, squares, or containers of signals for production control	0.43
Flow_01	Products are classified into groups with similar processing requirements	0.44
Flow_02	Products are classified into groups with similar routing requirements	0.45
Flow_03	Equipment is grouped to produce a continuous flow of families of products	0.53
Flow_04	Families of products determine our factory layout	0.48
Flow_05	Pace of production is directly linked with the rate of customer demand [a]	
Setup_01	Our employees practice setups to reduce the time required	0.59
Setup_02	We are working to lower setup times in our plant	0.45
Setup_03	We have low set up times of equipment in our plant	0.49
Setup_04	Long production cycle times prevent responding quickly to customer requests (reverse)[a]	
Setup_05	Long supply lead times prevent responding quickly to customer requests (reverse coded) [a]	
SPC_01	Large number of equipment / processes on shop floor are currently under SPC	0.48
SPC_02	Extensive use of statistical techniques to reduce process variance	0.52
SPC_03	Charts showing defect rates are used as tools on the shop-floor	0.59
SPC_04	We use fishbone type diagrams to identify causes of quality problems	0.52
SPC_05	We conduct process capability studies before product launch	0.61
Empinv_01	Shop-floor employees are key to problem solving teams	0.57
Empinv_02	Shop-floor employees drive suggestion programs	0.50
Empinv_03	Shop-floor employees lead product/process improvement efforts	0.58
Empinv_04	Shop-floor employees undergo cross functional training	0.62
TPM_01	We dedicate a portion of everyday to planned equipment maintenance related activities	0.42
TPM_02	We maintain all our equipment regularly	0.44
TPM_03	We maintain excellent records of all equipment maintenance related activities	0.47
TPM_04	We post equipment maintenance records on shop floor for active sharing with employees	0.42

[a]Eliminated during item purification in exploratory phase.

References

Ahmad, S., Schroeder, R.G., Sinha, K.K., 2003. The role of infrastructure practices in the effectiveness of JIT practices: implication for plant competitiveness. Journal of Engineering Technology Management 20 (3), 161–191.

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 103 (3), 411–423.

Anupindi, R., Chopra, S., Deshmukh, S.D., Van Mieghem, J.A., Zemel, E., 1999. Managing Business Process Flows. Upper Saddle River. Prentice Hall, NJ.

Boyer, K.K., Pagell, M., 2000. Measurement issues in empirical research: improving measures of operations strategy and advanced manufacturing strategy. Journal of Operations Management 18 (3), 361–374.

Boyer, K.K., Verma, R., 2000. Multiple raters in survey-based operations management research. Production and Operations Management 9 (2), 128–140.

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. Sociological Methods and Research 21 (2), 230–258.

Campbell, D.T., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait–multimethod matrix. Psychological Bulletin 56 (2), 81–105.

Cua, K.O., McKone, K.E., Schroeder, R.G., 2001. Relationships between implementation of TQM, JIT, and TPM and manufacturing performance. Journal of Operations Management 19 (2), 675–694.

Dankbaar, B., 1997. Lean production: denial, confirmation or extension of sociotechnical systems design? Human Relations 50 (3), 653–670.

Davy, J.A., White, R.E., Merritt, N.J., Gritzmacher, K., 1992. A derivation of the underlying constructs of just-in-time management systems. Academy of Management Journal 35 (3), 653–670.

Dean Jr., J.W., Bowen, D.E., 1994. Management theory and total quality: improving research and practice through theory development. Academy of Management Review 19 (3), 392–418.

De Treville, S., Antonakis, J., 2006. Could lean production job design be intrinsically motivating? Contextual, configurational and levels-of-analysis issue. Journal of Operations Management 24 (2), 99–123.

Devlin, S.J., Dong, H.K., Brown, M., 1993. Selecting a scale for measuring quality. Marketing Research 5 (30), 12–17.

Dillman, D.A., 2000. Mail and Internet Surveys: The Tailored Design Method, 2nd ed. Wiley, New York.

Doty, D.H., Glick, W.H., 1994. Typologies as a unique form of theory building: towards improved understanding and modeling. Academy of Management Review 19 (2), 230–251.

Dow, D., Samson, D., Ford, S., 1999. Exploding the myth: do all quality management practices contribute to superior quality performance? Production and Operations Management 8 (1), 1–27.

Fabrigar, L.R., McCallum, R.C., Wegener, D.T., Strahan, E.J., 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 4 (3), 272–299.

Finn, R.H., 1970. A note on estimating reliability of categorical data. Educational and Psychological Measurement 30, 71–76.

Filion, F.L., 1976. Estimating bias due to non-response in mail surveys. Public Opinion Quarterly 40, 482–492 (Winter 1975–1976).

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

Flynn, B.B., Sakakibara, S., Schroeder, R.G., 1995a. Relationship between JIT and TQM: practices and performance. Academy of Management Journal 38 (5), 1325–1360.

Flynn, B.B., Schroeder, R.G., Sakakibara, S., 1994. A framework for quality management research and an associated measurement instrument. Journal of Operations Management 11 (4), 339–366.

Flynn, B.B., Schroeder, R.G., Sakakibara, S., 1995b. The impact of quality management practices on performance and competitive advantage. Decision Sciences 26 (5), 659–691.

Hall, R.W., 1987. Attaining Manufacturing Excellence: Just in Time, Total Quality, Total People Involvement. The Dow Jones-Irwin/APICS Series in Production Management.

Hair Jr., J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, 5th ed. Prentice Hall, Englewood Cliffs, NJ.

Hopp, W.J., Spearman, M.L., 2004. To pull or not to pull: what is the question? Manufacturing and Service Operations Management 6 (2), 133–148.

Koufteros, X.A., 1999. Testing a model of pull production: a paradigm for manufacturing research using structural equation modeling. Journal of Operations Management 17 (4), 467–488.

Koufteros, X.A., Vonderembse, M.A., 1998. The impact of organizational structure on the level of JIT attainment. International Journal of Production Research 36 (10), 2863–2878.

Koufteros, X.A., Vonderembse, M.A., Doll, W.J., 1998. Developing measures of time based manufacturing. Journal of Operations Management 16 (1), 21–41.

Kozlowski, W.J., Hattrup, K., 1992. A disagreement about within group agreement: disentangling issues of consistency versus consensus. Journal of Applied Psychology 77 (2), 161–167.

Krafcik, J.F., 1988. Triumph of the lean production system. Sloan Management Review 30 (1), 41–52.

Li, S., Subba Rao, S., Ragu-Nathan, T.S., Ragu-Nathan, B., 2005. Development and validation of a measurement instrument for studying supply chain management practices. Journal of Operations Management 23 (6), 618–641.

Long, J.S., 1983. Covariance Structure Models: An Introduction to LISREL. Sage, Beverly Hill, CA.

Lorsch, J.W., 1980. Organizational design: a structural perspective. In: Koontz, H., O'Donnel, C., Weinrich, H. (Eds.), Management, a book of readings. McGraw Hill, NY, pp. 50–58.

MacCallum, R.C., Browne, M.W., Sugawara, H.M., 1996. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods 1 (1), 130–149.

MacDuffie, J.P., 1995. Human resource bundles and manufacturing performance: organizational logic and flexible production systems in the world auto industry. Industrial and Labor Relations Review 48 (2), 199–221.

McLachlin, R., 1997. Management initiatives and just-in-time manufacturing. Journal of Operations Management 15 (4), 271–292.

McKone, K.E., Weiss, E.N., 1999. Total productive maintenance: bridging the gap between practice and research. Production and Operations Management 7 (4), 335–351.

Meredith, J., 1993. Theory building through conceptual methods. International Journal of Operations & Production Management 13 (5), 3–11.

Meyer, A., Tsui, A., Hinings, C.R., 1993. Configurational approached to organizational analysis. Academy of Management Journal 36 (10), 1175–1195.

Miceli, M.P., Jung, I., Near, J.P., Greenberger, D.B., 1991. Predictors and outcomes of reactions to pay-for-performance plans. Journal of Applied Psychology 76, 508–521.

Miller, D., Friesen, P.H., 1978. Archetypes of strategy formulation. Management Science 24 (7), 921–933.

Monden, Y., 1981b. Adaptable Kanban system helps Toyota maintain just-in-time production. Industrial Engineering 13 (5), 29–46.

Monden, Y., 1983. Toyota Production System: A Practical Approach to Production Management. Industrial Engineers and Management Press, Norcross, GA.

Nahm, A.Y., Vonderembse, M.A., Koufteros, X.A., 2003. The impact of organizational structure on time-based manufacturing and plant performance. Journal of Operations Management 21 (3).

Narasimhan, R., Swink, M., Kim, S.W., 2006. Disentangling leanness and agility: an empirical investigation. Journal of Operations Management 24 (5), 440–457.

Ohno, T., 1988. Toyota Production System: Beyond Large Scale Production. Productivity Press, Cambridge, MA.

Osigweh, C.A.B., 1989. Concept fallibility in organizational science. Academy of Management Review 14 (4), 579–594.

Pil, F.K., MacDuffie, J.P., 1996. The adoption of high-involvement work practices. Industrial Relations 35 (3), 423–455.

Podaskoff, P.M., Organ, D.W., 1986. Self reports in organizational research: problems and prospects. Journal of Management 12, 531–544.

Priem, R., Butler, J., 2001. Tautology in the resource-based view and the implications of externally determined resource value: further comments. Academy of Management Review 26 (1), 57–66.

Putnam, H., 1978. Meaning and the Moral Sciences. Routledge& Kegan Paul, London.

Rinehart, J., Huxley, C., Robertson, D., 1997. Just Another Car Factory? Cornell University Press, Ithaca, NY.

Ross, J.E., 1993. Total Quality Management: Text, Cases, and Readings. St. Lucie Press, Delray Beach, FL.

Sakakibara, S., Flynn, B.B., Schroeder, R.G., 1993. A framework and measurement instrument for just-in-time manufacturing. Production and Operations Management 2 (3), 177–194.

Sakakibara, S., Flynn, B.B., Schroeder, R.G., Morris, W.T., 1997. The impact of just-in-time manufacturing and its infrastructure on manufacturing performance. Management Science 43 (9), 1246–1257.

Schmalensee, D.H., 2003. The perfect scale. Marketing Research 15 (3), 23–26.

Shah, R., Goldstein, S.M., 2006. Use of structural equation modeling in operations management research: looking back and forward. Journal of Operations Management 24 (1), 148 169.

Shah, R., Ward, P.T., 2003. Lean manufacturing: context, practice bundles, and performance. Journal of Operations Management 21 (2), 129–149.

Shrout, P.E., Fleis, J.L., 1979. Interclass correlations: uses in assessing rater reliability. Psychological Bulletin 86 (2), 420–428.

Sitkin, S., Sutcliffe, K.M., Schroeder, R.G., 1994. 'Distinguishing control from learning in TQM: a contingency perspective. Academy of Management Review 19 (3), 537–564.

Spear, S., Bowen, H.K., 1999. Decoding the DNA of the Toyota Production System. Harvard Business Review 77 (9/10), 97–106.

Sugimori, Y., Kusunoki, K., Cho, F., Uchikawa, S., 1977. Toyota Production System and Kanban system: materialization of just-in-time and respect-for-human system. International Journal of Production Research 15 (6), 553–564.

Wacker, J.G., 2004. A theory of formal conceptual definitions: developing theory building measurement instruments. Journal of Operations Management 23 (4), 629–650.

Whetten, D.A., 1989. What constitutes a theoretical contribution? Academy of Management Review 14 (4), 490–495.

Womack, J.P., Jones, D.T., 1996. Lean Thinking: Banish Waste and Create Wealth in Your Corporation. Simon & Schuster, New York.

Womack, J.P., Jones, D.T., Roos, D., 1990. The Machine that Changed the World. Harper Perennial, New York.

Wood, S.J., Stride, C.B., Wall, T.D., Clegg, C.W., 2004. Revisiting the use and effectiveness of modern management practices. Human Factors and Ergonomics 14 (4), 413–430.