Journal of Operations Management

WILEY

JOM FORUM OPEN ACCESS

Leveraging Strategic Demand Variability: The Role of Lead-Time Measurement

Rajan Suri[1] | Suzanne de Treville[2] | Tyson R. Browning[3] | Roger W. Schmenner[4]

[1]Department of Industrial and Systems Engineering, University of Wisconsin, Madison, Wisconsin, USA | [2]Faculty of Business and Economics, University of Lausanne and Swiss Finance Institute, Lausanne, Switzerland | [3]Neeley School of Business, Texas Christian University, Fort Worth, Texas, USA | [4]Kelley School of Business, Indiana University, Bloomington, Indiana, USA

Correspondence: Rajan Suri (rajan.suri@wisc.edu)

Received: 14 December 2023 | **Revised:** 10 June 2025 | **Accepted:** 18 June 2025

Handling Editors: Elliot Bendoly and Rogelio Oliva

1 | Introduction

The operations management field has long recognized that a key operations objective is to match supply with demand: Either an operation must be able to respond to demand variability, or demand must be stable. There was a strong sense in the late 1980s and early 1990s that responsiveness was an important element of competitiveness, leading to the exploration of time-based strategies (Stalk 1988; Schmenner 1988; Stalk and Hout 1990; Blackburn 1991). Tapping into this energy, I and colleagues enthusiastically embarked on ambitious projects to help interested companies increase their responsiveness by reducing their lead times so that they could profit from high-mix, low-volume, and customized products—only to realize that detailed principles for lead-time reduction had not been developed for this context. Although there was general agreement that reducing lead times was important, managers attempting to act on these insights struggled to prioritize reducing lead time if it also appeared to reduce utilization, efficiency, or increase costs in other ways.

In 1993 I founded the Center for Quick Response Manufacturing (QRM) at the University of Wisconsin-Madison as a university-industry partnership to address this challenge and help companies to gain competitiveness and profit from investing in lead-time reduction even when it appears to increase operating cost. Our objective over these decades has been to develop, implement, and test principles to provide guidance for such companies. The resulting tools and protocols—compiled into the QRM manufacturing strategy—have been widely applied over the ensuing three decades at hundreds of companies. The QRM Center has engaged in an average of 15 projects per year with its over 300 member companies, so over 400 company projects. It has organized around 10 workshops each year that have been attended by hundreds of managers, as well as regular international conferences every 1–2 years at which case studies of QRM implementation and their results have been presented. As a further development, in 2017 the QRM Institute, headquartered in Belgium, was established by an international network of organizations to provide access to QRM education, research, and diffusion of results in Europe. I serve as an advisor to the Board of the Institute. The initial QRM principles were summarized in a book entitled *Quick Response Manufacturing* that was published in 1998.

The previous co-EICs of *JOM* requested that I write a Forum piece for the *Journal of Operations Management* (*JOM*) to reflect on the experience and learnings from the QRM Center and QRM Institute from a *JOM* perspective. Exploring QRM from the viewpoint of *JOM* led me to the article by Schmenner and Swink (1998) that summarized their Theory of Swift, Even Flow (TSEF). Whereas QRM was developed to respond to the needs of companies, deploying the mathematical principles of lead time with a focus on hands-on implementation in a large number of companies, Schmenner and Swink (1998) took a philosophy of science approach to propose a theory to explain what makes an operation productive. Although the path to QRM is quite different from that which led to the TSEF, the key elements are remarkably similar. The exercise of examining the

https://doi.org/10.1002/joom.70005

QRM experience through a *JOM* lens provides some strong evidentiary support for the TSEF, demonstrating the usefulness of building bridges between perspectives.

2 | The Theory of Swift, Even Flow (TSEF) and QRM

The TSEF argues that the productivity of an operation rises with the speed and the evenness of the flow of work. The speed of flow depends on bottleneck capacity. It is improved when process steps that do not add value are eliminated, noting that production problems like defects and downtime create temporary bottlenecks. Throughput time is proposed by the theory as a way to measure flow speed in the process, starting from when work begins on a unit and ending when the unit is delivered or enters finished goods. Cellular manufacturing is proposed by the TSEF as a way to focus on a given type of product, building on the focused factory concept proposed by Skinner (1974) and expected to improve flow. Importantly, Schmenner and Swink (1998: 104) state:

> ...the Swift, Even Flow theory argues for the abandonment of numerous performance measures. Measures such as machine utilization or labor efficiency...are not measures of either flow or variability. For this reason, the theory argues that they should be abandoned as measures in favor of throughput time and variability (say, delivery performance to plan). Indeed, there is some confirmation that machine utilization and labor efficiency are not associated very much with productivity (Schmenner 1991; Schmenner and Vollmann 1994).

Up to this point, as I will soon show, the TSEF has a remarkably strong overlap with QRM, with each of the above points being key elements of the QRM approach. The only apparent disagreement concerns tolerance for demand variability. Schmenner and Swink (1998: 102) state:

> For materials to flow more evenly, one must narrow the variability associated with either the demand on the process or with the process's operations steps. Variability is measured by the variance or standard deviation of the timing or quantities demanded or of the time spent in various process steps. Variability is narrowed when the demands placed on the process are even and regular. 'Level' production plans are more compatible with productivity than are production plans with irregular quantities or due dates.

Yin et al. (2017) used the TSEF to explore the effectiveness of the *seru* production system in which assembly and testing are carried out in configurable cells. Small, general purpose equipment is mounted on wheels, allowing cells to be created in response to demand. *Seru* has been used in the electronics industry in Japan to permit manufacturing close to product development. This thus gives an example of creating a smooth, swift flow by deploying capacity so as to respond to demand variability. As predicted by the TSEF, productivity increased with flow—and this took place even though demand variability was mitigated rather than reduced.

2.1 | Improving Flow Speed When Demand Variability is High

The QRM Center focus, in contrast, has been on high-mix, low-volume products, often manufactured to order, and sometimes customized. Because this product profile does not lend itself to reduction of demand variability, a capacity buffer had to be installed at bottleneck operations in order to ensure a smooth flow. Demand variability can be a source of waste, but it can also be a source of profit and competitive advantage when it comes from meeting very specific customer needs that are only clearly defined soon before delivery. We then encouraged companies to divide variability according to whether or not it was a potential source of competitive advantage—whether or not it was *strategic* variability (Suri 2010).

Building a strategy around demand variability often involves trading off efficiency for speed, thus cost for time. Our experience continues to be that the companies that we work with struggle to reduce lead times when it requires that time be prioritized over cost, and this limits their ability to gain leverage and profit from strategic variability. Helping companies to manage this trade-off has thus been a top focus for the QRM Center.

2.2 | MCT: The QRM Time-Based Metric

Early in our journey, as we embarked on projects with numerous companies, it became clear that we would need to develop an appropriate time-based metric. We reached the same conclusion as Schmenner and Swink that delivery lead time—the time from when a customer places an order until when they receive it—would not be an appropriate metric for our purposes because it depends on the inventory that is in stock at the time of the order. We worked with our industry partners over several years to create a suitable metric with clear rules that we named Manufacturing Critical-path Time (MCT). MCT is defined as "the calendar time from when a customer submits an order, through the critical path, until the first end-item of that order is delivered to the customer." (Suri 2014). Our measure is aligned with the recommendation in the TSEF to use throughput time, although MCT starts from order submission rather than processing of the first piece. In other words, we count all the order processing steps as part of the throughput time given the make-to-order nature of the high-mix, low-volume products. With this in mind, we also developed supporting QRM principles for office operations (such as estimating, quoting, order entry, and engineering), material planning and scheduling, and supply management (Suri 1998).

Since the formalization of the MCT metric, all QRM implementations have been based on this metric. We do not claim that MCT is the optimal way to measure lead time, only that empirical evidence shows that it has been easy to use consistently and good enough to support QRM implementation. Our experience has been that deploying a consistent and precisely defined measurement approach has been essential if managers are to prioritize lead time in contexts in which it makes operations look less efficient. Also, the fact that all companies implementing QRM are making use of this measure makes it possible to compare results across companies. Although there are several precise and detailed rules to be observed in calculating MCT (see Suri 2014) we provide here an overview of the key concepts.

The MCT clock starts running in calendar days from the date of the order since that corresponds to how the customer is counting time. It follows the critical path of activities making up a product based on the following rules:

- All necessary activities must be completed from scratch. If, for example, components need to be fabricated for this specific order, then the MCT must include the time required for that fabrication. Components that are common to many products and are regularly held in stock do not need to be included in the MCT calculation; rules for this decision are also specified.
- All the processes needed for order fulfillment—order processing, materials planning, scheduling, manufacturing, and logistics—are included in the MCT.
- All the normal queuing, waiting, and moving delays that jobs incur are included.
- Time spent by material related to the product at any stage of the process—including that held at any inventory holding points—is included in MCT.
- MCT is based on the critical path for the product, that is, the longest path in manufacturing the unit.

As companies reduced their MCT significantly for a given set of products, they began to see substantial reductions in overall cost exactly as predicted by the TSEF. They also experienced quality improvements and major improvements in on-time delivery. The lessons learned, along with examples of key success stories, were documented in a second book (Suri 2010). [For readers not familiar with QRM, we recommend Suri (2010) as a starting point for understanding the details of this strategy]. In the following section, I describe how this played out in two QRM implementations.

3 | Company Examples

3.1 | RenewAire

RenewAire is a small manufacturer of customized energy recovery ventilation systems, based in Waunakee, Wisconsin. In 2003, RenewAire was struggling to match supply with demand. Daily fire-fighting was required to get orders out on time, and many orders were late. Customers placed orders well in advance of delivery according to the company's quoted lead time, and RenewAire released these jobs into production on schedule—which should have been early enough to deliver on schedule. Unfortunately, however, it often turned out that the customers got a better idea of what exactly was needed as the delivery date approached, so requested changes to their orders. An architect for a building being constructed, for example, would use early drawings to estimate the size of the RenewAire air exchanger to be installed at the time when the RenewAire order needed to be placed. But, as construction began, drawings frequently evolved in a way that had implications for recovery ventilation. One architect, for example, proposed replacing one large air exchanger by two smaller units because the end customer gained a better understanding of how air exchange would affect building comfort. This required updating the order even though manufacturing had begun on the original order. This extra work and inventory increased bottleneck utilization, which increased manufacturing lead times still further.

The president, Chuck Gates, had made an effort to reduce demand variability, but the fact was that the customer did not have sufficient information about what exactly was needed at the time they needed to place the order. And, it was the ability to provide the customer exactly what was needed that allowed RenewAire to compete profitably against much larger competitors offering a standard product. Interactions with the QRM Center led Mr. Gates to set as an objective the reduction of manufacturing lead time enough so that production could begin once the customer had a clear idea of their ventilation needs.

The first step in reducing the manufacturing lead time was to reorganize the shop floor into cells, which is the starting point for a QRM reorganization and aligns completely with the TSEF recommendation. These cells were designed around market segments, so that each cell was focused on meeting the needs of a given set of customers, consistent with Schmenner and Swink's (1998) observation that cellular manufacturing provides a path to factory focus.

Given that the objective was that each cell would have short enough manufacturing lead times to deliver on time even though production began after demand was known, management took steps to ensure that there was a 20% capacity buffer in each cell. While the capacity buffers did not need to be huge to ensure that waiting time at bottlenecks was minimized, they were large enough to make it look like efficiency on the shop floor was down and manufacturing cost was up. Mr. Gates therefore made the decision to replace efficiency and cost measures on the shop floor with MCT, such that MCT was the only measurement that was followed. This elimination of cost-based measures again follows precisely what the TSEF would have recommended.

In less than a year, RenewAire had reduced its MCT by over 80%. In less than 5 years, the company increased revenue by 140%, taking business away from much larger competitors. Productivity increased, such that the headcount in manufacturing increased by only 70% during this time. Mr. Gates shared these results and described the increase in profit at the company at a QRM conference in 2015.

Over two decades later, RenewAire continues to grow both by continuing to increase market share and by entering new markets. The small company that we started working with in 2003 is now considered an industry leader in energy recovery ventilation systems, and is known for its responsiveness. Mr. Gates retired several years ago, and a new president and vice president of operations are steering the company. Despite this change in leadership, RenewAire continues firmly on the same path. The new President, Scott Forest, expressed his ongoing support of decisions to reduce lead time, even if they incur extra costs: “Recently, I received a request to put a waterjet cutter in a cell that makes components for our ventilation systems. A traditional analysis showed that one machine would provide just enough capacity. However, since the components are critical parts made to order for each system, I didn't want our deliveries to be delayed if there was any backup in the components cell. So I actually approved the purchase of *two* machines, to provide spare capacity and flexibility to keep these components from holding up our deliveries. I recognize that our ability to provide customers with exactly what they need, in a very short lead time, is a huge competitive edge for us, and I regularly support that with my financial decisions.” Mike Ketter, Vice President of Operations, provides a key insight into how QRM has been sustained over the long term. “We put a lot of effort into making time-based thinking part of the fabric of our company. One technique we have used for this is to hold book clubs where, as part of their onboarding process, new employees read ‘It's About Time’ (Suri 2010) in small groups, and meet once a week to discuss one chapter at a time. To underscore top management's commitment to this initiative, I personally lead the group discussions each week, and then we award certificates to the participants in a ceremony after the last meeting. In my opinion, the book club has been key to getting our employees on board and making time-based thinking part of our organizational culture.”

As part of its QRM journey, Renewaire also implemented the full set of points that underlie the TSEF, and its results provide strong empirical support for the validity of the theory.

3.2 | Provan

Provan is a small Belgium-based metalworking subcontractor and supplier of metal products, offering a total solution for welded structures, laser and sheet-metal work, profile machining, and assembly. At the time Provan approached the QRM Institute in Belgium, the company had been combining an Enterprise Resource Planning (ERP) system with Lean practices to manage its order flow. The variation of parts being ordered had grown and batch sizes had shrunk considerably, and customers were requesting reductions in the time between order and delivery.

These factors resulted in excessive inventory and increased lead times. Ben Proesmans, one of the owners, decided to implement QRM company-wide after attending a workshop that I gave in Belgium. Like RenewAire, Provan organized manufacturing into cells around target markets, planned for enough spare capacity at bottlenecks to maintain flow, and made MCT their primary metric. Provan also implemented the card-based POLCA production control system on the shop floor, which is QRM's alternative to kanban for high-mix, low-volume, or custom production (Suri 2018). In less than a year, the MCT for a major line of products was reduced from around 4 weeks to 3 days—an 85% reduction—and on-time delivery performance went to 100%. The MCT of 3 days made it so that customers could place their order after knowing their demand.

As a result of Provan's performance, a major European customer re-sourced some of its products from a low-cost offshore provider to Provan even though Belgium has among the highest labor costs in the world (Pollet and Proesmans 2018). The customer calculated that Provan's extremely short lead times and high delivery reliability justified an 11% cost premium relative to the low-cost offshore provider because of cost savings with respect to items such as warehousing space and personnel, rush freight charges, and replanning and rescheduling time. Provan received the prestigious Factory of the Future Award from the Belgian government because of their contribution to Belgium manufacturing.

4 | Conclusions and the Way Forward

It has been a pleasure to reflect on how the more engineering and modeling-based approach developed in QRM aligns with the philosophy-of-science approach taken simultaneously by *JOM* researchers. We agree that the speed of flow in a production process depends on sufficient capacity. It is facilitated by organization of production into cells that are focused around product types. We also observed from our QRM company experiences and agree with TSEF that as speed of flow increases, so does productivity; hence prioritizing flow-based measures over measures of efficiency is likely to improve productivity more than a pure focus on efficiency.

Thus, the QRM experience provides strong evidentiary support for the theory that improving flow leads to improved productivity. We have also observed that companies that do everything except prioritize time-based measures over measures of efficiency have difficulty in justifying the capacity buffers at bottlenecks that are required to maintain flow.

QRM has developed in a high-mix, low-volume context, where demand variability may be a source of competitiveness and profit. We have also observed that local manufacturing in a high-labor-cost environment has been easier for companies to justify for high-mix, low-volume products. In these cases, the extra product changeovers add enough value to justify the increased capacity at bottleneck operations. The TSEF states that demand variability must be reduced in order to maintain an even flow. From the QRM experience, I suggest that this claim can be adjusted to state that demand variability must be sufficiently reduced *relative to the effective capacity of the cellular organization* to maintain a swift, even flow. This is consistent with, and complements, the findings in Yin et al. (2017).

Reflecting on this convergence leads me to claim that the path to swift, even flow remains ripe for further exploration. I hope that the QRM strategy and related tools are a source of ideas for bringing these effective concepts back to the forefront of manufacturing and operations strategy, and that our 32 years of improving flow and reaping productivity gains are useful to *JOM* researchers moving this powerful theory forward in practice.

5 | Commentary

Tyson R. Browning (t.browning@tcu.edu)

I greatly appreciate Suri's decades of experience developing, honing, and implementing QRM. It is essential that the *JOM* community incorporate insights such as these into our research and teaching. I find several of Suri's points particularly instructive in that regard.

I am a proponent of a value-oriented paradigm of operations management (e.g., Browning and de Treville 2021). Value means different things to various stakeholders. When customers value supplier responsiveness and short lead times, a competitive strategy that prioritizes them will indeed pay off. Here, *strategic variability* is an opportunity rather than a problem. While variability always adds costs, operations geared to handle it with the least increase in cost and when this increase leaves room for profit margin can capitalize on variability rather than avoid it (due to their relative advantage over competitors). QRM provides the DNA for operational systems that thrive in the seemingly harsh environment of variability. Thus, external variability need not always be deemed wasteful and sought to be tamed; it can be harnessed as a source of benefit instead. This shifts the strategic focus from efficiency to effectiveness as the key driver of value, and it provides a broader view of operational excellence than defect or waste reduction.

Of course, having the right strategy is necessary but insufficient for profitability. Operational capabilities must also align well with that strategy—and to provide real improvements, process changes must increase that fit. A key aspect of increasing this alignment is measuring what really matters, and QRM provides a case in point. Instead of measuring the utilization or efficiency of resources, focusing on the flow (or throughput) of what customers value reframes the problem. A focus on throughput/flow and proper measurements thereof was also a large part of Goldratt's insight in *The Goal* (Goldratt and Cox 2004) and the theory of constraints. This way of thinking continues to be counterintuitive and eye-opening for a great many students and managers. They find it difficult to fathom how reducing the productivity of part of a process (e.g., by decreasing its utilization) increases the productivity of the entire process. QRM provides another demonstration of this important principle.

A distinctive reframing in QRM is its focus on the MCT measure, which aggregates the time between order receipt and start of production with the production process flow time and the delivery (to customer) lead time. This change in primary metric shifts the focus from merely reducing process throughput/flow duration to also reducing (1) pre-process times, including those for order processing and backlog waiting, and (2) post-production delivery to the customer. This requires integration with and the improvement of other business processes (Browning 2020) and the ability to start work more quickly—which suggests an emphasis on production rate/capacity and cycle time (the average time between successive units) via the addition of capacity buffers for key activities. While Goldratt emphasized the importance of spare capacity (by definition) at non-bottleneck activities, to absorb internal variability and balance flow, Suri notes the importance of spare capacity at bottlenecks as well, to absorb external variability. Indeed, organizations likely need more than one type of "healthy fat" (Browning and de Treville 2021) to operate smoothly and consistently profitably. Moreover, by including the post-production delivery time, the "cost" of long transit times from production site to customer is also "counted"—an important, first step in highlighting its detriments.

These areas bring rich opportunities for empirical research. What operational strategies work well amidst the contemporary uncertainties and variabilities? Which operational capabilities and practices fit best with various strategies? What is the value of using the right metrics, and how much value is lost by using the wrong ones? How do different stakeholder foci, risk and value preferences, and time horizons affect strategic, operational, and measurement decisions? How have firms managed strategic transitions, such as to QRM, in terms of operational practices and metrics? How much are various customers willing to pay for increasing levels of responsiveness, and what are the tipping points for reshoring? These are just a few of the research questions that arise. I appreciate Suri exhibiting how the "right" operations and process improvements depend entirely on a firm's strategic direction, and that operational excellence can look quite different as a result.

6 | Commentary

Suzanne de Treville (suzanne.detreville@unil.ch)

In 1998, Schmenner and Swink (1998) published their article on TSEF. That same year, Suri published his book describing QRM. Coming from different directions, the frameworks presented in this literature are remarkably coherent. As Suri describes above, the TSEF/QRM principles have been shown to work. As long as flow takes priority—which requires that it be measured—productivity increases. The underlying insight emerging both from the theory-development approach published here in *JOM* and the engineering and first-principles-based approach explored in QRM was enthusiastically embraced as plausible by the greater Operations and Supply-Chain Management community. Importantly, this insight draws attention to the trade-off underlying how flow is achieved that exists between utilization and variability of demand and process. Flow can be achieved via a capacity buffer at bottlenecks, opening the door for high mix, low-volume, and customized products.

Three decades later, although time-based competition has had some successes (think Dell and Zara) I see this major realization as underutilized. During these decades, manufacturing over the developed world has shifted to countries offering a reduced per-unit landed cost. The loss of manufacturing jobs has been costly to local economic communities—although some communities have managed to reinvent themselves around a new industry, life in many others has dramatically worsened and remained that way. The loss of manufacturing know-how from chips to rare earths to shoes has had important consequences for the geopolitical environment. The assumption has been that developed economies are better positioned to offer services than manufacturing, relegating manufacturing to countries where the cost of labor is low, workers are willing to accept working conditions that would not be acceptable in the developed world, environmental protection is relatively limited, and energy is abundant. Glimpses of competitive manufacturing offered in the developed world from QRM experience and from the *seru* approach described by Yin et al. (2017) and cited by Suri offer a path to maintaining manufacturing close to demand and innovation, but have not led to widespread exploration.

The massive wave of offshoring that has occurred over these three decades has shifted the objective of manufacturing from building the product that the customer actually needs, to producing massive quantities of standard items that someone might eventually find attractive enough to purchase if the price is low enough. Apparel production provides an example: "Fast fashion" has allowed apparel manufacturers to respond to changes in consumer demand, but has also resulted in global overproduction with major environmental implications. The standard newsvendor model illustrates clearly that when the operating margin is high and the per-unit cost of overstock modest, the profit-maximizing order quantity may easily amount to several times median demand. The resulting inventory bloat fills stores, warehouses, and container ships, with much eventually ending up in landfill. Production is not to demand, it is to what is considered to be a potential bargain. And, the oversupply of products in some industries does not protect from undersupply in other industries when there are supply-chain disruptions.

The insights from QRM and TSEF together map out a path to producing to demand. Rather than ordering a container of shirts, phones, semiconductor chips, or pharmaceuticals, QRM/TSEF shows a path to producing the item that the customer wants to their specifications. QRM illustrates the competitive benefits that arise from meeting high mix, low-volume demand. This has the effect of transforming manufacturing into a service operation that happens to deliver a product. As clearly demonstrated by QRM and theoretically argued by the TSEF, this kind of manufacturing does not need to take the form of massive assembly lines with 20-s cycle times and poor working conditions, but can rather take the form of cellular manufacturing (whether in the fixed cells that have been successful in QRM, or the configurable assembly and test cells that define *seru*). The work conditions in such cells can be as attractive as work in more traditional service operations like those found in restaurants, hotels, retail, and health care.

Although the per-unit cost of local production to order may appear to be considerably higher than the per-unit landed cost from a low-cost supplier, a simple numerical example illustrates that the real options created by producing to order—postponing the decision about what to produce until demand is known—may represent more money than the apparent cost savings. Suppose, for example, that demand at the SKU level for some product (apparel, mobile phone, integrated circuit…) follows a lognormal distribution with volatility sigma = 0.8. Let's normalize price to 100, with the per-unit landed cost from the distant supplier only 10 per unit. A unit not sold during the demand period is scrapped. The newsvendor profit-maximizing service level is 90%, corresponding to $z = 1.28$ standard deviations above median demand. The resulting order from the distant supplier is thus $e^{z\sigma} = 2.79\times$ median demand. Expected demand is $e^{\sigma^2/2} = 1.38\times$ median demand. The fill rate for the proposed order is 89%.[1] This yields expected sales at 1.38*0.89 = 1.22× median demand, and leftover inventory at 2.79–1.22 = 1.57× median demand. The profit per unit of median demand is expected to be 94.32. Let's compare this order to placing an order for the same product from a local supplier once demand is known. We achieve the *same expected profit* if we pay the local supplier 31.5 per unit—a 215% premium relative to the low-cost supplier. And, this does not take into consideration other risks that are avoided (quality, logistics, customs problems) or follow-on options with respect to service or possible customization that are created. In paying the local supplier over 200% more per unit, we exchange a higher per-unit cost for a reduction in supply–demand mismatches.

The QRM/TSEF insights thus make clear that alternatives exist to filling container ships with a tsunami of cheap goods that get deposited into markets and end up in landfill. We can make everything for which there is actual demand, with total production cost comparable to that from the low-cost supplier. This supply chain does a better job of meeting demand, better supports innovation, permits creation of high-quality jobs, is likely to be profitable and competitive, dramatically reduces waste by only producing what is demanded, and redistributes production in a way that supports geopolitical stability.

7 | Commentary

Roger Schmenner (rschmenn@iu.edu)

Rajan Suri's *JOM* Forum article does a marvelous job of underscoring the principles that lead to improved productivity. His Center tackles what arguably is the hardest set of circumstances in which to foster productivity, that of high mix, low volume manufacturing. We should applaud all that they do.

I cannot help but look back at the over 50 years since I began studying factories in earnest. Progress has indeed been made, and that progress is consistent with the insights of QRM and the TSEF. Here are some reflections:

- Factories are tidier now and less crammed with inventory and conveyors going every which way.
- Layouts and the flow of work are more transparent.
- Fifty years ago, the quality movement, as exemplified by Deming and Juran, was just starting to catch hold. The crusade that was then being waged was over Material Requirements Planning. Both movements have won the day.
- Back then we talked of "purchasing" and pondered how much power over price the manufacturer had. Today, we talk of supply chains and ponder how best they can be coordinated.
- Large batch sizes and economies of scale were the rage. Today, we understand that speed and flexibility trump scale.
- Today, of course, automation is everywhere. Years ago, it was often reserved for the most tedious and unsafe steps in the process.
- The factories then most vulnerable to obsolescence from the incipient force of Japanese companies and their just-in-time manufacturing philosophy were hybrid processes that married batch operations to assembly lines. Today, that vulnerability is much attenuated. Lessons have been learned.

- Today, there is much lamenting about the transfer of manufacturing from places in the US where once it thrived. However, manufacturing has always been on the move. My earliest research was on industry location and it was clear then that companies persistently shifted production to lower cost locations. Consider the industrialization of the American South and the importance of right-to-work laws to location decisions. New England is a case in point of the need to consistently re-make oneself (e.g., abandoned textile mills, the demise of the mini-computer).
- The recent pandemic exposed the fragility of our current supply chains. But for me, it was simply a blip in the long, fruitful march that has cut out waste and cost from our manufacturing.
- Five decades ago, we taught Production. Now, of course, it is Operations Management. Services that now employ the bulk of the labor force have been usefully studied with the same principles that we applied to manufacturing.

It has only been a bit over a century since the moving assembly line revolutionized production. And it was only about 50 years before that when we saw the first continuous flow processes (e.g., oil refining). Those were step functions in the history of productivity. On reflection, the advances they represent are very consistent with the thinking that Professor Suri shares in his article. The vision of quick response/swift flow characterizes those achievements in business history. That same vision will inform our future.

Acknowledgments

Rajan Suri is deeply indebted to Tyson Browning and Suzanne de Treville for inviting this article, and for their time in providing outstanding advice on earlier versions of it. Suri would like to thank three anonymous referees whose constructive feedback helped to focus the article. Several people contributed to the original development and application of the MCT metric: they are acknowledged on page iii of Suri (2014). Suri would also like to acknowledge the contributions of Ananth Krishnamurthy and Charlene Yauch: Suri stepped down as QRM Center Director after the first 15 years and, as the subsequent Directors, they were responsible for conducting around half of the 400 company projects mentioned in this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Endnotes

[1] The fill rate is calculated as $\Phi(z-\sigma)+(1-\Phi(\sigma)) * e^{z\sigma-\sigma^2/2}$. For the derivation, see (de Treville et al. 2014).

References

Blackburn, J. D., ed. 1991. *Time-Based Competition: The Next Battle Ground in American Manufacturing*. Business One Irwin.

Browning, T. R. 2020. "Editorial: Operations Management Writ Large." *Journal of Operations Management* 66, no. 5: 494–500.

Browning, T. R., and S. de Treville. 2021. "Editorial: A Lean View of Lean." *Journal of Operations Management* 67, no. 5: 640–652.

de Treville, S., I. Bicer, V. Chavez-Demoulin, et al. 2014. "Valuing Lead Time." *Journal of Operations Management* 32, no. 6: 337–346.

Goldratt, E. M., and J. Cox. 2004. *The Goal*. 3rd Revised ed. North River Press.

Pollet, P., and B. Proesmans. 2018. "From Belgium: Metalworking Subcontractor Reduces Lead Times by 85% Using QRM and POLCA." In *The Practitioner's Guide to POLCA: The Production Control System for High-Mix, Low-Volume and Custom Products*. Productivity Press.

Schmenner, R. W. 1988. "The Merit of Making Things Fast." *Sloan Management Review, Fall* 1988: 11–17.

Schmenner, R. W. 1991. "International Factory Productivity Gains." *Journal of Operations Management* 10, no. 2: 229–254.

Schmenner, R. W., and M. L. Swink. 1998. "On Theory in Operations Management." *Journal of Operations Management* 17: 97–113.

Schmenner, R. W., and T. Vollmann. 1994. "Performance Measures: Gaps, False Alarms, and the Usual Suspects." *International Journal of Operations & Production Management* 14, no. 12: 58–69.

Skinner, W. 1974. "The Focused Factory." *Harvard Business Review* 52: 113–121.

Stalk, G., Jr. 1988. "Time: The Next Source of Competitive Advantage." *Harvard Business Review* 66, no. 4: 41–51.

Stalk, G., Jr., and T. M. Hout. 1990. *Competing Against Time: How Time-Based Competition Is Reshaping Global Markets*. Free Press.

Suri, R. 1998. *Quick Response Manufacturing: A Companywide Approach to Reducing Lead Times*. Productivity Press.

Suri, R. 2010. *It's About Time: The Competitive Advantage of Quick Response Manufacturing*. Productivity Press.

Suri, R. 2014. *MCT Quick Reference Guide*. C&M Printing.

Suri, R. 2018. *The Practitioner's Guide to POLCA: The Production Control System for High-Mix, Low-Volume and Custom Products*. Productivity Press.

Yin, Y., K. E. Stecke, M. Swink, and I. Kaku. 2017. "Lessons From *Seru* Production on Manufacturing Competitively in a High-Cost Environment." *Journal of Operations Management* 49–51: 67–76.