CHAPTER 8

Factory Performance Metrics: The Good, the Bad, and the Ugly

In my work, I routinely encounter firms—particularly those involved in the manufacture of high-tech products—in which data for literally hundreds of factory performance metrics are collected. Such efforts provide these firms with an enormous amount of raw data and keep a sizable workforce of automation personnel, information technology (IT) people, and (alas) industrial engineers occupied.

In most instances, however, the data—once processed—provide little, if any, information. More distressing, many of the metrics developed and acted on actually are counterproductive; that is, they encourage employment of factory protocols that worsen rather than improve factory performance—or else they support decisions that could have been accomplished more cost-effectively. In short, substantial resources—in terms of human energy and funding—are devoted to ill-advised and non-value-added efforts because of a reliance on inferior metrics.

In this chapter I first introduce a number of useful, fair, valid, and credible factory performance metrics (i.e., the good metrics). These include the

- Waddington effect plot
- M-ratio metric
- Availability profile plot
- Cycle-time contribution factor (CTCF)
- Degree of reentrancy (DoR) metric

Illustrations of how to gather the data and compute each of these metrics are provided in the material to follow.

This chapter also includes a discussion of some widely employed but problematic metrics (i.e., the "bad" and the "ugly"). Included among these are inventory or work-in-progress (WIP) turns, moves, utilization (of personnel or machines), and cost—as well as any non-load-adjusted factory velocity (e.g., cycle time and X-factor) metric or any measure derived by means of an inadequate data sampling rate.

The employment of these "bad" and "ugly" metrics may be worse than not measuring anything at all because they encourage and lead to poor decisions. Old habits, even those that harm a firm's bottom line, are, however, difficult to break.

The chapter concludes with a discussion of ways in which metrics—even "good" metrics—may be inflated or otherwise "gamed." In some firms, more time and resources may be devoted to gaming metrics than to improving factory performance.

WADDINGTON EFFECT AND PLOT

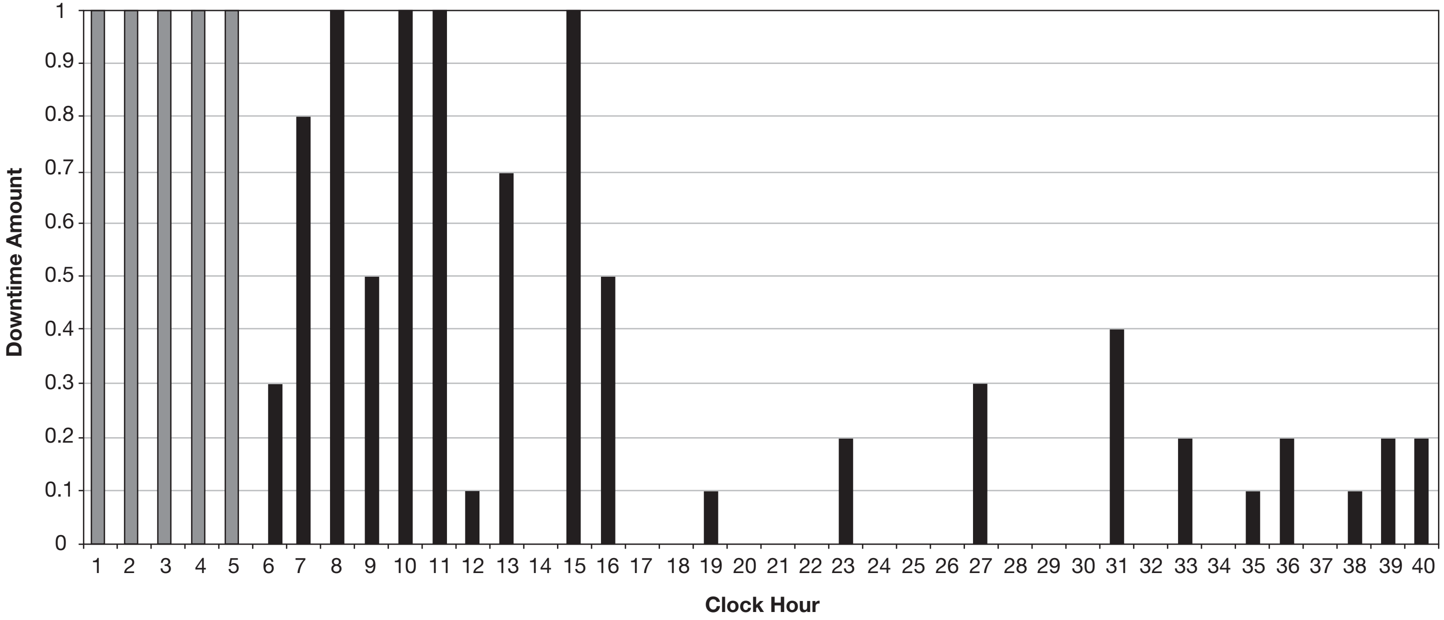
In Chapter 2, mention was given to the role that operational research played in World War II. One of the topics cited there was that of the Waddington effect, a phenomenon I named in honor of C. H. Waddington (Waddington, 1973). Waddington's operational research team, assigned to the British Coastal Air Command, encountered and identified the Waddington effect during the conduct of an effort to increase the flying hours of the Coastal Air Command's beleaguered air fleet.

In essence, it was found that the preventive maintenance (PM) events being employed appeared to induce rather than reduce unscheduled repairs. This disturbing effect was particularly evident "shortly" after the conduct of a PM event.

While Waddington's team dealt with a fleet of aircraft, their findings apply equally well to the "fleet" of machines in a factory. Figure 8.1 is a Waddington effect plot for a single-machine workstation in an actual factory. The workstation in question was scheduled to undergo a major PM event, requiring shutdown of the machine, every 40 hours. The PM event then in use required, on average, five hours to conduct (i.e., the time represented by the shaded bars in Figure 8.1). Following the PM event, the solid black bars indicate unscheduled downs (e.g., unanticipated repairs and/or unscheduled recalibrations) and their duration in hours

FIGURE 8.1

Waddington effect plot.



(see the vertical axis). For this machine, there were a total of five hours lost each 40-hour period to the major PM event and a further 9.2 hours lost to unscheduled downtime.

The Waddington effect may be recognized by an increase in unscheduled downtime *closely* following a PM event. In this illustration, a total of 6.4 hours of unscheduled downtime occurs within just 10 hours of completion of the PM event.[1] From then on, the profile of unscheduled downtimes decreases—until a similar pattern (not shown in the figure) is induced 40 hours later by the next major PM event.

Had the PM event accomplished its goal (i.e., to eliminate unscheduled downtime until the next PM event), the availability of this machine would be (we assume, for sake of discussion, that there is no blocked time)

$$A = \frac{(40 - 5)}{40} = 87.5 \text{ percent}$$

However, when the unscheduled downtime (9.2 hours in total) is included, the actual machine availability is just

$$A = \frac{(40 - 5 - 9.2)}{40} = 64.5 \text{ percent}$$

Furthermore, a full 16 percent of the machine's availability is lost to the unscheduled downtime in just the 10 hours following the PM event.

Waddington's group found that the unscheduled time lost closely following a PM event was a consequence primarily of poorly designed and/or poorly performed preventive maintenance (including scheduling PM events too frequently). The approach they developed (termed herein as a *Waddington analysis*) to eliminate or at least mitigate the Waddington effect will be described in Chapter 10.

To summarize, in an existing factory, data should be collected on the average time consumed by both scheduled (e.g., PM events) and unscheduled downtime (e.g., repairs and recalibrations) for

---

[1] The downtime in the 10 hours following the conduct of the 5-hour PM event includes 0.3 hour in hour 6, 0.8 hour in hour 7, 1 hour in hour 8, 0.5 hour in hour 9, 1 hour in hour 10, 1 hour in hour 11, 0.1 hour in hour 12, 0.7 hour in hour 13, 0 hours in hour 14, and 1 hour in hour 15 for a total of 6.4 hours.

both individual machines and workstations. These data should be plotted against clock time (e.g., as in Figure 8.1). If visual inspection (or automated pattern-recognition analysis) indicates the existence of the Waddington effect, further action (to be described in detail in Chapter 10) most definitely should be taken.

M-RATIO

The *M-ratio* (a.k.a. *maintenance ratio*) is the ratio of scheduled downtime to unscheduled downtime—and it may be computed for either individual machines, workstations, or an entire factory. The formula employed to determine the M-ratio is[2]

$$\text{M-ratio} = \frac{\text{scheduled downtime}}{\text{unscheduled downtime}} \tag{8.1}$$

Referring back to Figure 8.1, we see that the total unscheduled downtime over the period of interest (i.e., the 40 hours between PM events) was 9.2 hours, whereas the scheduled downtime was 5 hours. Assuming that these data represent averages typical of this workstation, its M-ratio is

$$\text{M-ratio} = \frac{5}{9.2} = 0.54$$

An M-ratio of 0.54 is in fact dreadful. It indicates that this workstation needs attention—and fast. For a typical factory (or machine or workstation), the M-ratio should be 9 or higher.

At an M-ratio of 9, the amount of unscheduled downtime is just 10 percent of the total downtime. In the preceding example, however, the amount of unscheduled downtime consumes 65 percent of the total downtime!

An M-ratio of less than 9 is usually an indication of a serious problem—most likely in the content, design, and implementation of PM events. Most often, however, factory engineers and managers would appear to want to believe that unscheduled downtime is caused by the design and/or operation of the machine, workstation, or factory (i.e., physical problems). While this may be the case, it is more likely that a low M-ratio is due to poor maintenance protocols.

---

[2] An alternative form of the M-ratio equation may be developed by including blocked time in the denominator of Equation (8.1).

AVAILABILITY PROFILE PLOT

An *availability profile plot* is used to record and analyze the availability of either a single machine or an entire workstation versus time. Figure 8.2 presents the availability profile plot of a hypothetical multiple-machine workstation. Samples of the workstation's availability (i.e., percentage of machines up, running, and qualified to support the processing responsibilities of the workstation) were taken every hour. While hypothetical, the plot is typical and similar in shape to those found in firms that fail to recognize the importance of protocols (particularly maintenance protocols).

Readers may note that the availability profile plot repeats itself every 12 hours, that is, every shift. Figure 8.3 presents a typical availability profile plot for just a single shift for this factory. A visual examination of the profile in either Figure 8.2 or Figure 8.3 reveals the fact that workstation availability is high (on the order of 90 percent) at the beginning of each shift and then plummets to a little more than 50 percent about two or three hours into the shift. About five or six hours into the shift, workstation availability has recovered and remains in the range of roughly 80 to 90 percent for the remainder of the shift.

FIGURE 8.2

A hypothetical availability profile plot.

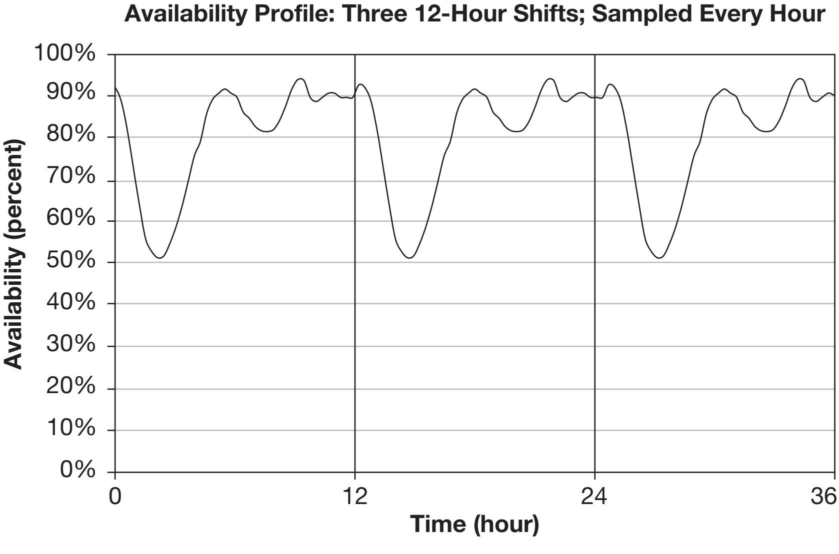
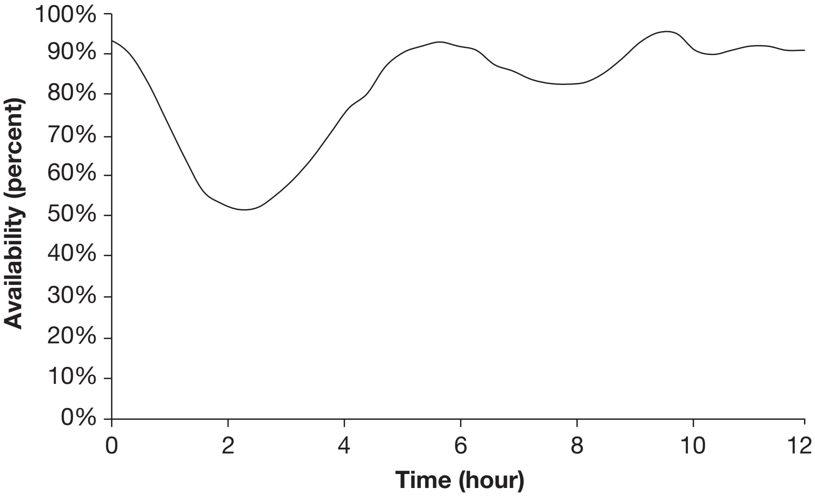


FIGURE 8.3

Availability profile plot, one shift.



We may compute the average workstation availability and its coefficient of availability from the data employed to plot the preceding figures. The results are

- Workstation average availability $A = 81$ percent.
- Workstation coefficient of variance of availability $CoV(A) = 0.164$.

Given these statistics and the availability profile plots, the next step in the analysis is—or should be—to ask ourselves why the profile takes on the shape exhibited in the figures. In fact, this was the question I asked myself many years ago when I first encountered an availability profile plot very similar to the one in Figure 8.2.

By observing operations on the factory floor, the reason for the pattern became evident. The factory workers had been warned to complete as many maintenance and repair events within their shift as possible rather than having such events extend across shifts. Consequently, floor personnel routinely scheduled as many maintenance events as doable as early into their shift as possible. This in itself was the reason for the shape taken by the availability profile.

While the factory manager, factory engineers, and floor personnel were satisfied with the results, the practice of clustering maintenance events (in this case, early in the shift) is a bad habit that degrades factory performance and should be avoided. Simply by means of declustering workstation maintenance events (and after overcoming significant resistance to this "radical recommendation"), a significant improvement in factory performance (in terms of both effective capacity and cycle time) was achieved.

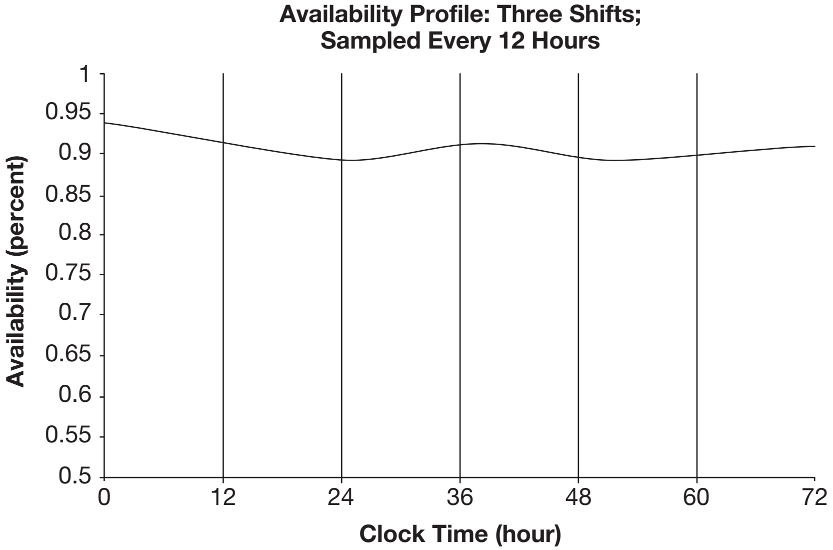
Before leaving the topic of availability profile plots, I must warn the reader that such plots can be—and too often have been—used improperly. Specifically, unless an adequate sampling rate is used, the plot is worthless. To illustrate, consider what would happen if instead of sampling availability every hour—as in Figures 8.2 and 8.3—we had sampled approximately every 12 hours (i.e., once per shift). Figure 8.4 shows the result of such a reduced sampling rate.

A visual examination of the new plot would suggest that workstation availability remains in a range between about 90 and 93 percent. The impression given is that this workstation is running very efficiently, whereas nothing could be farther from the truth.

We also may compute the average workstation availability and its coefficient of availability from the data employed to plot Figure 8.4. The results are

FIGURE 8.4

Availability profile plot, samples taken every 12 hours.



- Workstation average availability $A = 91$ percent, in contrast to 81 percent when an adequate sampling rate is used.
- Workstation coefficient of variability of availability $CoV(A) = 0.013$, in contrast to 0.164 when an adequate sampling rate is used.

Again, the impression is given of an efficient, stable workstation when, in fact, its actual performance is, at best, problematic. To summarize, unless the sampling rate employed to gather data is adequate, conclusions contrary to reality may be drawn.

CYCLE-TIME CONTRIBUTION FACTOR

The *cycle-time contribution factor (CTCF),* while not infallible, provides a reasonably effective means to assign priorities to the dedication of resources for the improvement of factory performance. More specifically, this particular metric may be used to help decide how to best allocate resources among process steps so as to improve overall factory performance.

In Chapter 3, the *cycle-time efficiency (CTE)* metric was defined. Recall that the *CTE* of a given process step is found by means of Equation (8.2):

$$CTE_{ps} = \frac{PT_{ps}}{CT_{ps}} \tag{8.2}$$

where PT_{ps} is the average raw process time of the process step, and CT_{ps} is the actual average cycle time of the step.

Consider, for example, the cycle-time efficiencies of two process steps at a specific level of factory loading. Assume for the moment that step *X* has a *CTE* of 10 percent, whereas that of process step *Y* is 20 percent. One may ask the question: To which of these two steps should priority be given to the dedication of resources for factory performance improvement at the existing level of factory loading?

While the *CTE* of process step *X* is less (and thus worse) than that of process step *Y,* it could be the case that dedicating resources to the improvement of the performance of step *Y* (e.g., by means of reducing the variability inherent in that step, by increasing the availability of the machines supporting that step, or by increasing their process rates) actually may be more beneficial to overall factory performance. This may be determined by means of computing the cycle-time contribution factor of the process steps in question.

To demonstrate, examine the reentrant factory of Figure 8.5. There are four workstations, each of which consists of one or more identical machines. The process flow is depicted in the figure. The degree of reentrancy (DoR) of the factory is 7/4, or 1.75.

We will assume that we know the raw process time (*RPT*) of each of the seven process steps and that we have expended the resources necessary to collect credible data on the average actual cycle times of these steps. These two sets of data are all that are required to determine the *CTCF* for each of the workstations' process steps.

The data are listed in Table 8.1 in the columns labeled "*PT*" and "*CT*," respectively. We then use Equation (8.2) to derive the

FIGURE 8.5

Four-workstation reentrant factory.

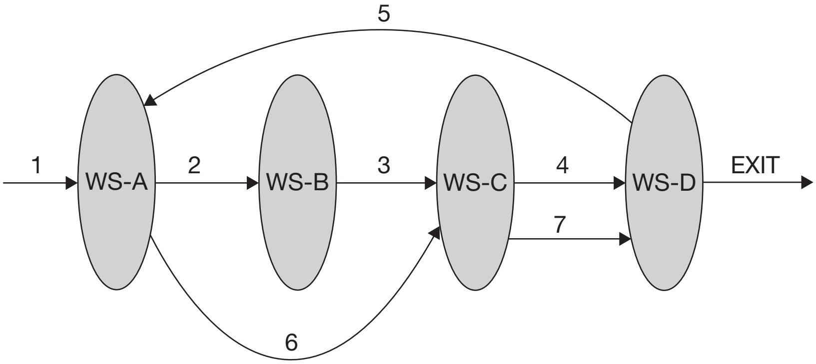


TABLE 8.1

Process step cycle-time contribution factors

Workstation	Process Step	PT	CT	CTE	CTCF
A	1	1	6	**0.167**	2.833
B	2	2	7	0.286	2.429
C	3	3	9	0.333	1.889
D	4	2	7	0.286	2.429
A	5	4	13	0.308	**1.308**
C	6	3	9	0.333	1.889
D	7	2	8	0.250	2.125

cycle-time efficiencies for each process step, as listed in the column headed by *"CTE."*

The final step in derivation of the cycle-time contribution factors is to employ Equation (8.3) for their derivation, where

$$CTCF_{ps} = \frac{\sum_{ps=1}^{P} PT_{ps}}{CT_{ps}} \tag{8.3}$$

For example, the *CTCF* for process step 5 is found by dividing 17 (the sum of all raw process times for all process steps) by the *CT* of the process step (i.e., 13).

It may be noted that although process step 1 has the worst *CTE* value (i.e., lowest *CTE*), the process step whose improvement likely will have the most impact on total factory cycle time is step 5 (i.e., the step having the lowest value of *CTCF*).

To repeat, the process step having the lowest *CTCF* value typically is the highest-priority workstation in terms of the allocation of resources. In this instance, process step 5 of workstation A has the highest priority, followed by process steps 3, 6, 7, 2, 4, and 1 in that order. Thus, although process step 1 has the worst *CTE,* it likely has less impact on overall factory cycle time (again, for the specific factory loading level under consideration) than process step 5.

Given that process step 5 has the highest priority, just some of the actions that might be taken to increase the value of its *CTCF* include

- Reduce the coefficient of variability of arrivals at process step 5.
- Increase the availability of the machines in workstation A that support process step 5.
- Reduce the coefficient of variability of the effective process times of the machines that support process step 5 (e.g., this might be accomplished by reducing the variability of the repair times of these machines).
- Increase the M-ratio of the machines that support process step 5 (e.g., this might be achieved via the conduct of a Waddington analysis).

Note that each of these actions should reduce the cycle time of process step 5 and thus increase the value of its *CTCF.*

DEGREES OF REENTRANCY

Readers were first exposed to the degree of reentrancy (DoR) metric in Chapter 3. As discussed, the DoR metric provides an indication of the complexity, as induced by reentrancy, of either a complete factory or any of its reentrant nests.

The ideal factory (approached to some degree by the Arsenal of Venice, the Ford Model T plant, or the Toyota production system) has a DoR of 1, that is, no reentrancy. Reentrancy in itself may, and usually does, induce variability—and it always increases complexity. One has only to observe the frantic (and mostly counterproductive) measures taken by engineers in a highly reentrant factory—say, a semiconductor wafer fabrication facility—in their attempts to dispatch jobs (a.k.a. *WIP management*) to reentrant workstations to appreciate the problems imposed by reentrancy.

Reentrant factories lead to countless mostly pointless and fruitless meetings in which the sole topic is that of the presentation of impassioned arguments for and against various WIP management schemes. Some personnel argue for "back-to-front" WIP management, some for "round-robin" (i.e., cyclic) dispatching, and others for the employment of "critical-ratio" WIP management, whereas others demand dynamic (as opposed to static) dispatching. The fact is, however, that given a reentrant or any other type facility, the only WIP management scheme that routinely and reliably improves overall factory performance is one that reduces the combination of batch forming times and departure-rate variability. This assertion should be obvious from a close examination of the second fundamental equation of manufacturing.

The optimal approach to the matter of WIP management is to focus on three factors: (1) reduce factory DoR, (2) reduce batch forming times, and (3) reduce the departure-rate variability from each process step. Subsequent chapters will deal with practical means to accomplish each of these goals.

SOME "BAD" AND "UGLY" PERFORMANCE METRICS

To qualify as "bad" and "ugly," a performance metric must be one that can and likely will lead to decisions that only worsen overall factory performance. Among the very worst of these are inventory (or WIP) turns, moves, and utilization, as well as (1) any non-load-adjusted factory velocity measure and (2) any metric derived from data collected at an inadequate sampling rate. Yet it is these measures, rather than the more valid and useful ones described previously [e.g., load-adjusted cycle-time efficiency (LACTE), factory operation curve, factory profit curve, Waddington effect plot, M-ratio, availability profile plot, cycle-time contribution factor, and DoR] that seem to be, alas, among the most commonly employed in actual practice. It should not be surprising, then, that the vast majority of factories (or supply chains and business processes) operate far below their potential.

Changing a factory manager's or engineer's mind, when it comes to these bad and ugly metrics, may prove to be an exceptionally difficult task. Hopefully, the discussion that follows will explain just why one should avoid (or, at the very least, appreciate the deficiencies of) the bad and ugly metrics. I begin with the metric known as *inventory* (or *WIP*) *turns.*

Inventory and WIP Turns

Inventory turns (or, alternately, WIP turns) are usually found by dividing average factory throughput *TH* by average factory inventory *WIP.* Thus one version of the formula for inventory turns is expressed as

$$\text{Inventory turns} = TH/WIP \tag{8.4}$$

However, since Little's equation states that $WIP = CT \cdot TH$, Equation (8.4) may be transformed into

$$\text{Inventory turns} = TH/WIP = TH/(CT \cdot TH) = 1/CT \tag{8.5}$$

Clearly, the shorter the cycle time (i.e., the faster factory velocity), the higher—and supposedly better—is the value of inventory turns.

An alternative representation of inventory turns, and one popular in the semiconductor manufacturing sector, is that of WIP turns, as given by

$$\text{WIP turns} = \frac{\text{number of value-added process steps}}{\text{cycle time}} \tag{8.6}$$

For example, if there are 150 value-added process steps in the manufacture of a product and the average cycle time is 75 days, then the WIP turns are 150/75, or 2.

A few of the problems with using WIP turns include the following:

- It may be difficult to impossible to come to agreement as to the number of value-added process steps (e.g., which steps truly add value and which don't?).
- Some factories may produce products having inherently longer raw process times. This increases, at no fault of the factory, the minimum possible cycle time.
- Cycle time is a non-load-adjusted metric. Simply by just slightly reducing factory loading, one factory may, for example, be able to halve its cycle time.

Inventory turns, because they employ non-load-adjusted cycle time (Equation 8.2) in their computation, suffer from the same problem as noted in the second and third bullets in the preceding list.

Moves

There are factory managers who use moves as a performance metric. A number of them would seem to even rely on moves as their primary means for assessing performance. They evidently believe that the greater the number of moves (e.g., jobs processed per unit time through the factory, workstation, or machine or by a human), the better the entity in question is performing.

To explain why moves are such a poor and misleading metric, all one has to do is to consider a hypothetical factory having a single and fixed constraint, no loss owing to scrap, no reentrancy, and no variability whatsoever. Assume further that the constraint workstation has an upper bound on capacity (i.e., maximum throughput rate) of 5,000 jobs per week. Thus, if 4,000 jobs per week are entered into this factory, the average number of jobs out (i.e., moves per factory per week) would be 4,000. If the jobs started were 4,900 per week, the average number of moves per factory per week would be 4,900. A starts rate of 5,000 jobs per week would, in this perfect factory, produce 5,000 moves per factory per week.

So far, all is well and good. If, however, we increase the jobs rate to any level above 5,000 per week, the number of factory moves will remain at 5,000. Even if we increased jobs introduced into the factory to, say, 20,000 jobs per week, we still achieve just 5,000 moves—a value dictated by the factory constraint. Of course, with any job starts rate above 5,000, the factory inventory level would increase—ultimately to infinity. Correspondingly, the average job cycle time would increase to infinity. Not many customers would be pleased with that amount of lead time.

The employment of moves as a performance metric encourages factory overstarts (i.e., the introduction of jobs started at a rate higher than the maximum sustainable rate of factory throughput). If, however, management is ignorant or dismissive of the value and importance of reduced factory cycle time, moves actually may be employed as the preferred performance metric.

In short, moves are a poor measure of performance because their use encourages bad behavior and ignores factory cycle time and factory inventory buildup. The situation is even worse for a more realistic factory, that is, one having variability, multiple and migrating constraints, reentrancy, and losses owing to scrap.

Utilization

Whether we are talking about the utilization of machines or human beings, utilization is an unfortunate and counterproductive measure of performance. When a factory manager sets some level of utilization as a (or the) factory goal, he or she has fallen into the trap of relying on perception rather than reality.

Factories in which utilization is a primary performance goal encourage the "look busy" syndrome. When a factory manager or a corporate executive walks through such a factory, workers scramble to find something—anything—to do that will give them the appearance of being busy. I have observed, on more than one occasion, maintenance technicians actually stopping perfectly good machines and performing unnecessary and unscheduled maintenance—simply to appear to be doing something. In other instances, I have seen workers snatch jobs that had successfully finished processing on their workstation and reintroduce them (i.e., reprocess them) into the workstation. Perhaps the cleverest approach I have observed was that of a machine operator who hid a stockpile of unfinished jobs next to his machine so as to avoid ever having his machine idle in the event of a management walkthrough. There seems to be no end to ways in which to convince either a moves-or utilization-obsessed manager that a worker or workstation is busy.

Utilization is the first cousin of moves. Both metrics encourage the worst behavior of workers, and both do more harm than good.

Cost

The final member of the gang of "bad" and "ugly" metrics to be discussed is cost. Unfortunately, despite its problems and negative impact on decisions, cost is often the primary metric used by management (and Wall Street analysts) to measure a factory's or firm's performance. Just as with moves and utilization, however, an emphasis on minimizing cost encourages—and rewards—bad behavior and poor decision making.

As discussed in Chapter 7, the primary goal of a for-profit firm should be to maximize profit and market share and to do so over the long rather than the short term. Firms that emphasize cost minimization often do so while ignoring profit and market share—or the health of the company—over the long term. Unfortunately, the reduction of cost by a firm (e.g., via layoff of employees, consolidation of operations, outsourcing, etc.) is almost always both encouraged and well received by Wall Street.

Not only is corporate management encouraged to cut costs to achieve some short-term benefit, but also so are the firm's employees. Just one illustration of this was evident in a firm whose CEO demanded cost cutting as well as purchasing delays so as to avoid having to report a loss for the quarter. The CEO did not want to be the firm's first chief executive to have to report a quarterly loss, and he made his wishes clear to management.

In order to satisfy the CEO's short-term (and, as it proved to be, short-sighted) goal, management at all levels demanded that purchases of new machines be put off for one quarter and notified employees that they would be rewarded if they found ways to reduce the number of machines in each factory. Purchases were indeed delayed—resulting in increased factory cycle times and the displeasure of the firm's customers. Factory engineers, eager to satisfy the machine-reduction goal, conducted detailed (and mostly flawed) analyses to determine if any machines could be removed (and either sold for salvage or simply shut down for some period of time) without affecting factory capacity. One engineer discovered that a few machines could be removed from two of the workstations in the factory without any evident impact on capacity. A few months later, this engineer was recognized in a meeting of all factory personnel—and presented with a $20 coupon that could be used (within 30 days) at a local restaurant.

The primary results achieved by this firm's cost-reduction effort were as follows:

- Costs for the quarter were indeed reduced, and the firm avoided having to report a quarterly loss.
- Factory cycle times increased significantly.
- Lead times promised to the firm's customers could not be satisfied.
- Customer dissatisfaction increased significantly.
- Effective (as opposed to predicted) factory capacity decreased somewhat.
- Shortly after the "successful" cost-reduction effort, two of the firm's biggest customers switched either all or part of their business to the firm's major competitor.

While one would imagine that in a sane and rational world, this particular cost-cutting exercise would have taught this firm a lesson, cost reduction apparently remains—at least at this point in time—the firm's major concern. Cost reduction is still rewarded, the impact and importance of factory cycle time has yet to be fathomed, and factory managers remain convinced that there is no need to be knowledgeable about the science of manufacturing.

GAMES PEOPLE PLAY

It is all well and good to point out good, bad, and ugly factory performance metrics. Unfortunately, however, even the best metric may prove to be ugly—or even hideous—if management fails to insist on accountability and oversight in the collection, processing, and interpretation of the underlying data. A hands-off manager, particularly one who refuses to gain a tolerable level of appreciation of the science of manufacturing, invariably will fall victim to bogus information. To illustrate, I will now discuss just a few of the ways in which metrics may be (have been and—alas—are being) gamed.

Gaming LACTE

Load-adjusted cycle-time efficiency (LACTE), one of the three holistic factory performance measures described in Chapter 7, is—when employed and interpreted properly—–one of the very best ways, if not *the* best way, to measure and compare factory performance. However, it is relatively easy to game this metric (this is true of any metric). Some of the approaches employed to misuse the LACTE metric include

- Ignoring the fact that a LACTE curve must be developed and using, instead, just a *LACTE* point value
- Intentionally or unintentionally understating actual factory capacity (e.g., assuming or pretending that the existing throughput of a factory employing inadequate or improper protocols represents the true, sustainable limit of its capacity)
- Intentionally or unintentionally inflating the value of the factory's raw process time

Gaming the Waddington Effect Plot and M-Ratio Metric

At one brief moment in time I had convinced myself that two metrics that could not be gamed were the Waddington effect plot and the M-ratio. However, I had underestimated the degree to which some individuals would go to produce bogus results.

In one firm, the engineers at one of its factories initially reported a very poor (i.e., low) M-ratio for their workstations. Once they discovered that their M-ratio performance (as, however, being computed correctly) was the worst in the firm, they found a way to triple the value of their M-ratio overnight. They simply omitted data produced by their poorest performing workstations, asserting that these were "outliers." The revised and bogus M-ratio value was cited in their next report. Other factories discovered the scheme and, in order to remain competitive, followed suit. Within a few weeks, all the factories in the firm had M-ratios anywhere from three to four times their true value. Management, ignorant of the practices being employed, congratulated everyone—and themselves—for achieving such excellent results in such a short time.

A similar rationale was employed subsequently in the development of factory Waddington effect plots. Unscheduled downtimes closely following a PM event were reclassified. Instead of being designated as unscheduled downs, many of these events were given such names as "pseudo-PM events" and "extended PM events." Thus, in one fell swoop—by means of self-deception—the Waddington effect was mitigated, and M-ratios were increased. The perpetrators of these practices, once again, were congratulated and rewarded for their "good work."

Gaming Cost

Since cost is such a popular metric, countless ways have been developed to report bogus cost reductions. Shifting expenditures into the future was discussed previously and remains a classic way to satisfy short-term cost-reduction goals.

With the proliferation of various management fads, some rather clever approaches to attributing cost reductions to the success of the fad *du jour* have been introduced. As just one illustration, a large multinational firm adopted a certain management cost-reduction fad [for the sake of discussion, let's call it *utopian management* (UM)] a decade or so ago. The firm's CEO stated that he wanted to see a reduction of at least 30 percent in costs as a consequence of the introduction of UM.

Not only was the 30 percent reduction goal achieved, but it was also surpassed. The cost reduction alleged to be achieved by the introduction of UM was on the order of 35 percent. This was accomplished, however, mostly by means of accounting trickery. Specifically, any cost reductions in the firm—achieved by any means—were attributed to the introduction of UM. For example, when one factory was shut down solely owing to obsolescence, UM was given the credit for the resulting reduction in cost. And when a factory engineer invented a new method for producing one of the firm's products, UM was given credit for the cost reduction.

As a result of the widespread reporting of the firm's alleged but mostly inflated success with UM, other firms eagerly adopted the fad. To their surprise, their results were not nearly so impressive. Such disappointments, however, have yet to slow the proliferation of articles, books, seminars, and courses on this particular fad.

Gaming Moves and Utilization

The metrics of moves and utilization are likely the most frequently abused measures of performance. The "look busy" culture that these metrics induce has been discussed.

Another way in which the moves metric is frequently abused is through the "cherry picking" of jobs to be introduced into a workstation. If management is imprudent enough to rely on moves, factory floor personnel soon will discover that given the choice between introducing a job that will take a long process time versus one that takes a short time, selecting the job requiring the shorter time will increase the number of moves.

The utilization metric may be and often is gamed by means of the introduction of unnecessary bureaucracy and red tape. At one firm, the approval of a minor, if not trivial change in the amount of supplies ordered for a workstation required the signatures of five different people. The requestor, a factory floor maintenance worker, hand carried the approval forms. The average time spent to obtain all five signatures was on the order of six hours—time that the requestor should have spent on the factory floor in support of the numerous unscheduled repair events the workstation encountered. Under this system, the average utilization of the workstation personnel was on the order of 80 percent.

This particular business process ultimately was changed—despite the resistance of floor personnel, who did not want to be seen as "less utilized." A clever but naive new hire on the factory floor conceived and developed an improved business process for the ordering of supplies. Despite the objections (and veiled threats) of his coworkers, he presented the concept to his supervisor. The method was approved and implemented soon thereafter. The new process, requiring a single approval—via e-mail—reduced the average approval time from six hours to less than an hour. The average availability of the associated workstation, as a consequence, increased from 74 to 86 percent. At the same time, the utilization of the workstation's personnel dropped significantly (i.e., the worker who had been spending much of his time gathering signatures was able to devote that time to the workstation).

Unfortunately, this story does not have a happy ending. About six months after introduction of the streamlined supply-ordering process, a "lean manufacturing" task force recorded the utilization of factory floor personnel. They found that the utilization of the floor personnel supporting the workstation in question was "only" about 60 percent—well below the 70 percent goal established by the firm's finance department. Failing to appreciate that lean manufacturing is, or should be, focused on more than cost reduction, the task force recommended a reduction in the number of personnel assigned to the workstation. As a consequence, the utilization of the workstation's floor personnel increased to more than 80 percent, and workstation availability decreased to roughly 75 percent. Factory cycle time took a similar hit. Perhaps even worse, the result confirmed the worst fears of the factory workforce, that is, that any improvements in business processes would be punished by reductions in workforce size and an increase in workload.

Gaming Is Widespread

Some readers may wonder if the practice of gaming is as insidious and widespread as I have implied. Perhaps, you may think, the preceding discussion has exaggerated such practices. Unfortunately, the gaming of performance metrics is all too common, whether in a factory, a mutual fund, a bank, or even the ranking of universities. As one example, in 2008, the *U.S. News & World Report* considered changing the way it ranks U.S. law schools (A. Efrati, "Law School Rankings May Change to Deter 'Gaming,'" *Wall Street Journal*, August 26, 2008, p. A1). The change considered attempts to deter a "popular practice" employed by law schools that involves channeling lower-scoring full-time applicants into part-time programs that don't count in the rankings.

The fact of the matter is that no matter what may be the results of any type of ranking (e.g., of factories, schools, or places to live), take these with a grain of salt. Unless there is a serious system for auditing and objectively analyzing such rankings, games can and will be played.

CHAPTER SUMMARY

If a factory performance metric is to be useful, it must satisfy certain conditions. These include

- The data employed to support the metric must be collected at an adequate sampling rate.
- If different factories are to be compared fairly or the changes in a factory's performance are to be evaluated properly, the metrics employed absolutely and positively must be load-adjusted.
- Discipline, accountability, and oversight (by those capable of doing so) are a necessity. Any performance metric, whether "good," "bad," or "ugly," is susceptible to gaming.

Factory performance metrics that do not satisfy these prerequisites encourage poor decisions and lead to degraded rather than improved performance.

My personal recommendations for performance metrics are

- Use holistic measures:
  - LACTE
  - Profit curve

- Use supporting measures:
  - Waddington effect plot
  - M-ratio
  - Availability profile plot
  - Degree of reentrancy

This said, I am convinced that far too much time and emotion are wasted on arguments concerning performance measures. Firms that most quickly achieve significant and sustained factory performance improvement focus the bulk of their efforts on simply reducing variability and complexity.

CASE STUDY 8: INTRIGUE IN THE PARTS WAREHOUSE

It's Saturday morning, and Winston is busy at work in his "war room." Sometime later, Julia and Dan arrive, announcing that Brad had told them that he might not be able to make it this morning.

"I do believe that Brad Simmons has something better to do this weekend," says Julia, grinning.

"What do you mean by that?" asks Dan. "What's up with Brad?"

"All I know," Julia replies, "is that I saw Brad and Sally Swindel having dinner at the Golden Goose last night. I've also been told that Sally asked her firm to transfer her to Hyperbola's local office—an office located about two blocks from here. Someone, I think, is going to be seeing a lot more of Sally Swindel."

"Wow," says Dan. "We could have a problem."

"I don't think so," Julia replies. "Brad's a good guy. If being with Sally Swindel makes him happy, then I'm happy for him."

Winston Smith clears his throat. "Brad may not be a concern, but we do have a problem. Walk with me to my cubicle, and you'll see what I mean."

Once the three are at Winston's cubicle, he points to another cubicle, located across the hall. "When I arrived at work yesterday, I noticed that someone had set up a cubicle, the one you see there. A few minutes later, a young man arrived and took a seat in the cubicle. He gave me a wave, I waved back, and that was the extent of our pleasantries. All day long, however, I caught him watching me. I don't think he's doing anything else. People, I think he's spying on me. We need to be very careful. Very careful."

* * *

Donna Garcia is also working this Saturday morning. She sits at the desk in her home office perusing the printouts of the e-mails that Winston Smith had sent to Tommy Jenkins—the same e-mails that were intercepted by Ben Arnold and had been considered to be of no value. Her efforts are interrupted by a phone call.

"Hello," says Donna. "Yes, I received your e-mail. Yes, it does seem that our Winston Smith has had a long relationship with Professor Leonidas. I've also found out that the professor was the person responsible for the turnaround of ToraXpress. Using his approach, they were able to reduce their factory cycle time by about 75 percent—without having to purchase any new machines. Somehow, they even increased their capacity. Then Muddle bought the company, forced our 'No Deviations' policy on them, and their performance turned rotten."

Donna pauses to listen to the other person on the line and then replies, "Don't worry. I've got someone watching Winston. I also had IT remotely download everything on his computer. We may find something there. In the meantime, I noticed that he sent e-mails to Tommy requesting approval to access the simulation software Muddle is using. He even asked if he could join forces with Muddle's simulation group. In one of the e-mails he claims that he is confident that we could reduce our factory cycle time by 60 percent or more."

Donna again pauses to listen. "Don't worry, Tommy is not going to know about any of this. Right now all the poor sap wants to do is to have the lowest cycle time of any of our factories."

* * *

Shortly after Donna Garcia hangs up her phone, Brad Simmons—on the other side of town—dials Sally Swindel's cell phone number.

CHAPTER 8 EXERCISES

1. The Waddington effect plots for two workstations in a factory indicate the existence of the effect for workstation A and no evidence of the effect for workstation B. Their

average availabilities, however, are identical. Why should one be concerned with workstation A?

2. Return to Exercise 4 of Chapter 3 and compute the M-ratio for the machine described.
3. An availability profile plot of a workstation in a factory indicates that its average availability is in the range of 85 to 90 percent for the night shift but lies in the range of 40 to 80 percent for the day shift. What might be the cause of this difference? What should be done?
4. A firm samples the availability of its workstations at the end of each shift (the value recorded is the average availability of each workstation over the entire shift). What argument would you present to the factory manager that might convince him or her to be wary of the results presented via such a practice?
5. List some of the ways in which the following performance metrics might be gamed. Try to come up with tricks not mentioned in this chapter.
    - M-ratio
    - Utilization
    - Availability profile plot
    - Factory operating curves

CHAPTER 9

A Transition: From Words to Deeds

The purpose of the preceding eight chapters was to present you with an introduction to and appreciation of the fundamentals that serve to determine factory performance. An awareness of factors that do and don't directly affect factory cycle time, capacity, and lead time is an essential first step toward taking the actions necessary to improve the performance of a factory or, for that matter, a business process or organization.

Now that this initial step has been taken, we may transit from history, terminology, equations, and concepts to a discussion of pragmatic and cost-effective means to achieve significant and sustainable factory performance enhancement in the real world. Simply put, the purpose of this brief chapter is to provide a transition from words to deeds.

The actions required do not necessarily stem from current philosophies or fashions (e.g., lean manufacturing, Six Sigma, or reengineering) or slogans or buzzwords. Rather, they are the tools of implementation—formed via a combination of experience and science—necessary to most effectively achieve the desired results.

The section that follows provides a brief recapitulation of the material covered in the preceding eight chapters. This is followed by a similarly brief introduction to the approach required to move from words to deeds—topics that will be addressed in detail in Chapters 10 through 15.

CHAPTERS 1 THROUGH 8: A LOOK BACK

One must never forget that the three enemies of factory performance, in terms of operations and maintenance (i.e., the actual running of the facility), are complexity, variability, and lackluster leadership. Attempting to improve performance by means of addressing topics other than these is almost always a waste of time and resources.

While there is much to be said for the notions encompassed in such efforts as lean manufacturing, Six Sigma, and the Toyota production system, and while—if applied properly—positive results may be achieved, their names themselves may get in the way of attaining significant and sustainable results. What, for example, does it really mean to achieve a "lean" factory? Depending on the conscious or subconscious intent of management, "getting lean" may mean anything from reducing the workforce to reducing cost to increasing utilization. These, however, are goals—and not necessarily the right goals. Goals, even if selected properly, are meaningless without (1) an understanding and appreciation of the factors that determine performance and (2) a practical plan (i.e., means) for achieving the goals.

While the title of this book, *Optimizing Factory Performance: Cost-Effective Ways to Achieve Significant and Sustainable Improvement,* is not intended to be cute, clever, or sexy, it summarizes precisely what is required for factory performance improvement in the complex and perplexing environment of the real world. Specifically, we must both appreciate and exploit the third dimension of manufacturing. Furthermore, three factors (i.e., politics, art, and science) must be considered if improved production-line operation is to be achieved and sustained. Moreover, if factory performance improvement is to be accomplished, it must be done in an intelligent (as opposed to a strictly emotional) and systematic manner.

In addition—as revealed in the 12-workstation problem of Chapters 4 and 6—without an appreciation of the three fundamental equations of manufacturing, one must substitute experience, judgment, guesses, hunches, intuition, speculation, and luck in place of science. The three fundamental equations, on the other hand, indicate precisely what factors determine capacity, cycle time, and the propagation of variability. Most important, they encourage a focus on actions that most likely will improve performance while avoiding less effective or even counterproductive decisions based on hunches and intuition.

While the forms of the three fundamental equations illustrated previously are appropriate only for very simple factories (e.g., those in which there is an absence of reentrancy, job-to-machine dedication, batching, and cascading), they may be and have been extended to more realistic and complex facilities. Such extensions are mostly of academic interest, however, and one need only comprehend the most basic forms of the three equations (i.e., those presented in Chapter 5) to gain the necessary appreciation of how factory performance is determined. For example, it now should be clear that

- A reduction of variability (e.g., of job arrivals, process times, down events, or wait times) always improves all aspects of factory performance.
- An increase in capacity (e.g., via faster process rates or increases in availability) may or may not improve all aspects of factory performance.
- Exploitation of the third dimension of manufacturing (i.e., changes in manufacturing protocols) provides a means to improve performance that is generally faster, cheaper, and more sustainable than decisions confined to the first or second dimension (i.e., physical changes).
- While balanced production lines running at the takt rate (i.e., the fundamental premise of lean manufacturing) may be appropriate for synchronous facilities (e.g., bottling plants and automotive assembly lines), they are often inappropriate for certain modern-day factories (i.e., asynchronous production lines, such as found in semiconductor manufacturing) (Stecke and Solberg, 1985).
- The theory of constraints provides an interesting and insightful supposition but has limited utility in real-world factories, where there exist multiple migrating constraints operating within a dynamic environment of variability and change.
- Complexity induces variability, which, in turn, degrades performance.
- Any reduction in complexity improves production-line performance and reduces stress on the workforce.
- An ideal factory should eliminate or at least reduce batching and cascading, incorporate the simplest production line possible (e.g., minimize DoR), and keep the

size of each machine ideally not more than four times the size of the job (or lot or batch) it processes—unless otherwise dictated by the laws of physics.

- The most overlooked means of improving a firm's bottom line is through reduction of factory, business process, and supply-chain cycle time.
- Since a factory is a nonlinear dynamic system with feedback, one's intuition is almost always wrong.
- Absent the support and involvement of management (up to and including the firm's CEO), any effort directed toward performance improvement is less likely to be successful and almost certainly will not be sustainable.

We also have discovered that many factory performance measures do not, in fact, do a particularly good job in objectively, fairly, and accurately assessing performance and, in fact, are likely to be counterproductive. Furthermore, organizations too often waste vast amounts of time and resources in never-ending arguments as to which metrics to use, how to collect the data, how to present the data, and even—alas—how to game the results.

In my experience, firms that have achieved real and lasting factory performance improvement devoted the bulk of their efforts to directly and effectively addressing the three enemies of performance rather than in purposeless and seemingly endless meetings. In short, there is far too much time wasted on words and far too little devoted to action.

The following section presents a brief introduction to the real-world implementation of the art and science of manufacturing. This discussion will be illustrated and elaborated on in the chapters that form the remainder of this book.

CHAPTERS 10 THROUGH 15: A LOOK FORWARD

The fundamental equations of manufacturing and (1) a familiarity with the history of manufacturing, (2) an acquaintance with the most effective measures of factory performance, and (3) an appreciation of the scope and limitations of both the art and science of manufacturing provide a solid foundation for the selection and implementation of the actions necessary to most quickly and effectively improve production-line performance. These actions, in turn, consist of those dedicated to

- The reduction of complexity
- The reduction of variability
- A more accurate determination of workstation and factory capacity
- Realization of the leadership necessary to overcome the political obstacles faced by any effort that involves change

Chapter 10 presents guidelines and illustrations of the most effective means to reduce the complexity of the protocols employed within the factory. Coverage includes the means to achieve a reduction in

- The number of process steps
- The factory's degrees of reentrancy
- The complexity (and ambiguity) of maintenance and operations specifications
- Clutter and confusion in the workspace
- Complexity of workstation run rules (a.k.a. *dispatch rules*)

Reduction of the sources and impact of variability is the focus of Chapter 11. The chapter provides illustrations and discussions of the reduction of variability in such areas as

- Variability induced through the clustering of
  - Factory starts
  - Maintenance activities
- Variability induced by the inefficient assignments of personnel to tasks (e.g., the allocation of maintenance personnel to workstations) and subsequent wait times
- Variability induced by the inefficient location and stocking of spare parts and supplies
- Variability induced by (invariably futile) attempts to chase work-in-progress (WIP) bubbles (i.e., the perils of reactive as opposed to proactive decision making)
- Variability induced by inappropriate job-to-machine dispatch (WIP management) rules (Ignizio, 2003b)
- Variability induced by batching and cascading operations
- Variability induced by the failure of management to act expediently and consistently and to provide the means and guidance necessary to manage change

Chapter 12 presents a simple example—using a modification of the original 12-workstation factory model—that serves to illustrate the way in which one may use the material in previous chapters to improve factory performance significantly. Rather than mathematical methods and models, only the guidelines for complexity and variability reduction need be employed. In essence, this chapter serves to put together, in one illustration, most of the ideas presented in previous chapters. For many readers, this may prove to be the most useful chapter in this book.

Chapter 13 provides an overview of the methods and models that should be employed to determine the true upper bound of workstation (and factory) capacity. When combined with the discussion of variability presented in Chapters 5 and 11, this chapter allows you to more accurately predict both maximum theoretical and sustainable capacity. The models and methods employed also may be extended to encompass, among other matters, the analysis of optimized job-machine dispatch rules and personnel-to-task assignments.

If the implementation approaches presented in Chapters 10 through 13 are to be successful, there must be a vision, a plan, a capable (and preferably experienced) team, the means to implement the plan, and first-rate leadership. In addition, consideration must be given to the politics of the organization and, in particular, any resistance to change. Chapter 14 deals with these topics.

Finally, Chapter 15 recalls the attributes of the ideal factory and provides a current example of a firm that appears to have incorporated, independently, the main concepts outlined in this book. This Spanish firm has turned conventional wisdom on its head and in so doing has achieved a corporate cycle time (i.e., time from the inception of a new product to its delivery to retail outlets) that is 95 percent less than that of its best-in-class competition.

CHAPTER SUMMARY

Having completed Chapters 1 through 8, you have attained the background necessary for practical, real-world, cost-effective factory performance improvement. You are now prepared to address the plans, decisions, and modes of implementation necessary to overcome the three enemies of factory performance—and to do so in the real world.

If only the Muddle Company would follow these guidelines!

CASE STUDY 9: EVERYBODY'S DOING IT!

Some six weeks after his unfortunate meeting with Marvin Muddle, Tommy Jenkins is presented with the latest performance numbers for Muddle's factories. A nervous pair of toadies waits as the facts sink in.

"What the heck is going on?" asks Tommy. "Our factory's cycle time decreased by 20 percent, but in the same time period Jack Gibson's factory's cycle time—Factory 2—dropped by 30 percent! I thought you two promised me that our cycle time would be the best in the firm by now. And just look at Factory 2's per-unit cost—that's been reduced by 20 percent!"

Donna replies, "Tommy, somehow Jack Gibson found out about our plan to reduce factory starts. He claimed—or lied, just as we did—that his factory capacity was lower than previously estimated and received permission to reduce his starts. That's how he achieved the reduction in cycle time. He also did something else. He ordered his people to cut costs by reducing their inventory of parts and supplies. They've pushed out their purchase requests until the next quarter. He also stopped funding activities like retirement parties, retirement gifts, and team building. He's even curtailed virtually all travel. Boss, he's really making us look bad."

"Two can play at that game," says Tommy. "Ben, send out an order to our department managers. Tell them that I want our expenses cut by at least 30 percent. In addition, I want to announce a 'Hurry Up' program. I want to see everyone on the factory floor busy all the time. And you, Donna, find out who leaked our plans to reduce factory starts. I want that person drawn and quartered!"

"You've got it, boss," Donna replies, barely suppressing a giggle.

Behind Tommy's back, Ben gives her the "thumbs up" sign.

* * *

It's 9 p.m., and we find Julia, Dan, and Winston busy in the "war room." Dan has noticed that the coolness between Julia and Winston has lessened. Right now they are sitting next to one another examining the results of the latest simulation exercise. Dan's attention is diverted by a message on his cell phone.

"Winston," says Julia, "just look at what's been achieved by no more than a reallocation of maintenance technicians. My goodness, the simulation shows that our factory cycle time could be reduced by a third."

"Not only that," adds Winston, "there's an opportunity to reduce maintenance personnel by at least 10 percent and still maintain the same improved cycle time. I bet that Marvin Muddle would love that. The man's obsessed with cost cutting."

"I hate to burst your bubble," says Dan, addressing the pair, "but I just received an e-mail on my BlackBerry from Professor Leonidas. It seems that Donna Garcia has asked him to teach her about the science of manufacturing. She offered to pay him $10 an hour for his services."

"Good grief," says Julia, "that's not much more than minimum wage. Can she be serious? Did the professor accept her 'generous' offer?"

"No," says Dan. "He said that he thanked her for her interest but just didn't have the time. She didn't take that well."

"This is strange," says Winston. "Why has Donna Garcia taken such a sudden interest in the science of manufacturing? She's about the last person in this company who I think would be interested in anything having to do with science."

"Something is up, that's for sure," says Julia.

"There's more," adds Dan. "Donna expressed a lot of interest in reducing cycle time and in factory operating curves. She also asked his opinion on simulation via fluid models."

"Something is fishy," says Winston. "I sent some e-mails to Tommy Jenkins about fluid models months ago. Like all the other e-mails, I got no response. Now, all of a sudden, one of Tommy's toadies is asking Aristotle about fluid models."

"Do we have," asks Dan, "anyone we can trust? Someone who might provide us with some information? We seem to be treading on dangerous ground."

"There is one person," says Julia. "Bridget Wallace is Ben Arnold's administrative assistant. Bridget also happens to be Brad Simmon's cousin. I know for a fact that she despises Ben. If it weren't for the money and medical coverage, she'd be long gone. We need to see if Brad can convince her to keep her eyes open."

"I'm not sure about that option," says Winston. "Brad seems preoccupied with this Sally Swindel person. Besides, have you ever given any thought to the fact that Brad may have gone over to the dark side? Perhaps your friend is the one leaking information to Donna Garcia."

"I don't believe that for a minute," Dan replies angrily, "Brad's not that kind of guy."

"I agree," says Julia. "Brad Simmons is a decent and honorable person. There's no way he would be trying to hurt us."

"But," says Winston, "if you recall, Brad wondered out loud as to what all our work here would do for us. As I recall, he implied that this is a waste of time, that someone else will take credit for our work."

This page intentionally left blank

CHAPTER 10

Reducing Complexity

The pioneering figures of scientific management, industrial engineering, and operational research emphasized, in their work, the reduction of complexity in the operation and control of factories, supply chains, business processes, and entire organizations. This focus also was evident in the efforts of the more enlightened nineteenth- and early-twentieth-century industrial firms—particularly the Ford Motor Company of the early 1900s.

As discussed in earlier chapters, the Toyota production system (a.k.a. *lean manufacturing*) systematized, revised, and enhanced many of these contributions. Toyota put together a unified package of numerous concepts and methods that ultimately led them to a world-class status in automobile manufacturing. Unfortunately, "waste walks" (i.e., identifying sources of waste by means of visits to the factory floor), "CANDO" (i.e., *c*leaning up, *a*rranging, *n*eatness, *d*iscipline, *o*ngoing improvement, a.k.a. *5S* or *workplace organization*), and "process-step reduction" form, for many firms, the main—and alas, sometimes sole—thrust of their attempts to introduce lean manufacturing.

When conducted by the right people with the right training in the right manner on the right problem, waste walks, CANDO, and process-step reduction can and usually do lead to performance improvement—at least over the short run. Such efforts represent, however, only a limited subset of the undertakings necessary to deal most effectively with all sources of factory complexity. The more insidious forms of waste escape detection when lean is limited to just these approaches.

The following sections provide discussions and illustrations of complexity-reduction efforts. The chapter concludes with an example that illustrates the implementation of a Waddington analysis, a tried, true, practical, and cost-effective methodology that reduces or eliminates most of the typical sources of complexity, be they in a factory or, as to be described, for improvement of the operating and maintenance protocols of the legendary "Brown Bess" musket (Antil, 2006; Wikipedia, Brown Bess Musket). We begin with the task of reducing unnecessary process steps.

PROCESS STEPS

A common thread of most efforts directed toward the reduction of complexity is their empirical, as opposed to scientific, basis. The reduction of complexity in the process steps of a production line is no exception. More specifically, complexity reduction typically is an art enhanced by experience and practice. Some individuals ultimately are able to become experts—true artists—whereas others, regardless of their training and actual experience, never progress beyond the ordinary.

Whatever one's skill level and natural gifts, the first step in the reduction of process steps is to develop a process flowchart, preferably a process-step-centric flow path. Recall that the development of such plots was illustrated in Chapter 3.

Unfortunately, it is rare to find a firm that has developed flowcharts at the level required for effective process-step reduction. This means that the organization or some outside consultant must construct the necessary flowcharts. The steps recommended for establishing a process-step-centric flowchart include the following:

1. Locate the individual or individuals in the firm who have detailed knowledge of and demonstrable expertise in the production-line process flow.[1]
2. Combine visits to the factory floor with carefully structured discussions with the firm's process flow experts. Use these visits and discussions to establish a preliminary, mutually agreed upon process-step-centric flowchart.

---

[1] Guidance on how to identify and work with experts and on the task of knowledge acquisition in general is provided in Chapter 5 of Ignizio (1991). One way to identify a domain expert (e.g., in the domain of process-step flow) is to determine the domain-expert bottleneck, that is, the person or persons who are busiest fielding questions and assisting in solving problems in the domain of interest.

3. Once satisfied that the process-step-centric flowchart reflects the true nature of the process flow, identify and list the steps that are the most likely candidates for either elimination (i.e., outright removal) or refinement (i.e., reduction in the complexity of the individual process step). Agreement on this matter should be accomplished jointly with the firm's process-step experts. The most common process steps for reduction or refinement are
   - Inspection steps
   - Inspection sampling rates
   - Transit steps (i.e., movement from one process step to the next)
   - Unnecessarily complex or superfluous process steps of any type
4. Evaluate the consequences of removal or refinement of the process steps identified in step 3. Ideally, this should be accomplished by exercise of a credible and validated simulation model. Absent the existence of such a model, the most effective alternative for evaluation is by means of a carefully planned and controlled pilot study on the actual factory floor.

Based on readings, discussions, and personal experience, I have noted that anywhere from 10 to 40 percent of the process steps employed by a typical firm may be eliminated or refined. One reason for unnecessary steps is the fact that it is not atypical to encounter a production line employing the same or nearly the same process-step flow as originally implemented (possibly years or even decades ago) for a given product. This is particularly true for inspection steps.

Typically, the initial inspection steps (and associated sampling rates) employed for a new product are conservative, that is, likely excessive. Fear, timidity, and unwarranted concern may serve to defer any thought of changes in these steps. Alternately, the potential improvement possible by changing the inspection steps simply may not be recognized. Once, however, confidence in the machines and methods employed to support a given step has been established, action (in the form of an analysis and evaluation of the consequences) should be taken to remove or refine the associated inspection step. Such an effort almost always will pay for itself many times over.

The reduction in the complexity of transit steps typically involves determining shorter, less complex routings between nontransit process steps. This may be extended to substituting transit via automated means (e.g., robots, conveyors, or monorails) for manual transportation (e.g., hand carries or movement by pushcarts). It should be noted, however, that the automation of transit may not necessarily produce less complexity or even improve performance. In fact, there have been instances in which a change to automated transit actually increased complexity, cost, and factory cycle time.

The ultimate impact of the reduction or refinement of process steps is determined very much by the parameters and configuration of the production line in question. Reduction in factory cycle times of anywhere from 5 to 20 percent appear to be typical, however. In at least one instance I am familiar with, process-step reduction and refinement reduced factory cycle time by more than a third.

DEGREES OF REENTRANCY

As discussed previously, reentrancy induces complexity (which, in turn, increases variability) and thus the ideal factory should have a degree of reentrancy (DoR) of 1. Engineers at the Ford Motor Company achieved significant and sustainable improvements in the cycle time for the Model T automobile “simply” by transforming their original reentrant flow path into a serial nonreentrant path. This was accomplished by locating each of the machines supporting a given process step according to the sequence of process steps.

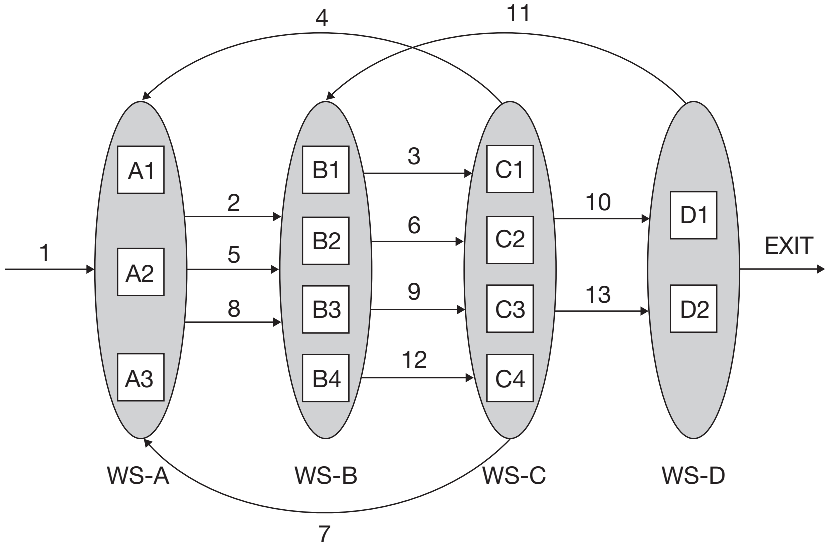
Factory engineers and plant managers often are hesitant to follow Ford’s example. Their intuition (and remember my caution about intuition) evidently tells them that this approach to reducing reentrancy will degrade rather than improve performance.

Much of this misplaced concern is based on the assumption that the more machines in a given workstation, the better is its (and the factory’s) performance. It seems obvious that the more machines in a workstation, the greater is its availability and flexibility. This assumption, however, may be invalid when the workstation supports reentrancy. This may be demonstrated by means of the following example.

Figure 10.1 depicts a four-workstation factory with a DoR of 3.25 (i.e., 13 operations supported by four workstations). The rate of arrivals at this factory (i.e., arriving at workstation A) is 1.5 jobs

FIGURE 10.1

Four-workstation factory (DoR = 3.25).



per hour. Transit time between operations will be assumed, for sake of simplicity, to be zero. We also will assume that there is no rework or scrap, and thus, if the factory is to be effective (and not allow the buildup of an unsustainable length queue), its output also should be 1.5 jobs per hour. The maximum theoretical capacity of each machine in each workstation is listed in Table 10.1. (The availability of all the machines in the factory has been assumed to be 90 percent.)

Note in Table 10.1 that the effective process rate (*EPR)* for process step 1 by any machine in workstation A is 1.5 jobs per hour, whereas the *EPR* for process step 7 by any machine in workstation A is 1.636 jobs per hour. Wherever blank spaces occur in the table (e.g., the machines in either workstation B, C, or D for process step 1), we interpret this to mean that those workstations do not support that particular process step.

While Table 10.1 lists maximum theoretical capacities by machine type and process step, Table 10.2 summarizes the maximum theoretical capacity of each workstation. It also serves to list the specific process steps supported by each workstation, the composite arrival rate of jobs at each workstation, and the workstation occupation rate ρ.

TABLE 10.1

Machine Maximum Theoretical Capacities

Effective Process Rates per Machine Type (Lots/Hour)

Process Step	A Machines	B Machines	C Machines	D Machines
1	1.5	—	—	—
2	—	1.5	—	—
3	—	—	1.5	—
4	1.5	—	—	—
5	—	1.636	—	—
6	—	—	1.8	—
7	1.636	—	—	—
8	—	1.5	—	—
9	—	—	1.5	—
10	—	—	—	1.5
11	—	1.5	—	—
12	—	—	1.5	—
13	—	—	—	1.8

TABLE 10.2

Workstation Maximum Theoretical Capacities EPR_{ws}

Workstation	Process Steps Supported	Composite Arrival Rate (Lots/Hour)	Workstation Theoretical Capacity (*EPR* in Lots/Hour)	Workstation Occupation Rate ρ
A	1, 4, 7	4.5	4.628248	0.972290
B	2, 5, 8, 11	6.0	6.127341	0.979220
C	3, 6, 9, 12	6.0	6.260870	0.958330
D	10, 13	3.0	3.272727	0.916670

The derivation of the values listed in Table 10.2 may be explained as follows:

- Composite arrival rate = the sum of the arrival rates of the jobs to be processed. For example, workstation A processes three steps (1, 4, and 7), and the rate of flow of each of these is 1.5 jobs per hour. Thus the composite arrival rate at workstation A is 3 • 1.5 = 4.5 jobs per hour.
- Workstation (maximum) theoretical capacity = the product of the number of machines in the workstation and the

harmonic mean of the effective process rates for every operation of each machine in the workstation. For example, the maximum theoretical capacity of workstation A is given by multiplying 3 (machines) by the harmonic mean of 1.5, 1.5, and 1.636. The spreadsheet representation of this is 3*Harmean(1.5, 1.5, 1.636), with a result of 4.628248. (Note that simply adding the individual *EPR* values of the machines in a workstation, as used in the 12-workstation demonstration, is appropriate only if the workstation supports a single operation, whereas the harmonic mean is appropriate if the workstation supports multiple operations, and every machine in the workstation has identical machine-to-operation process rates. For more general situations, the methods of Chapter 13 must be employed.)

- Workstation occupation rate = the ratio of the workstation's composite arrival rate and its maximum theoretical capacity. The occupation rate of workstation C, for example, is given by 6 divided by 6.26087, resulting in $\rho = 0.95833$.

Since the occupation rate of each workstation is less than 1, it is (at least theoretically) possible that the factory can support the process flow. Whether this is practical or not (i.e., in terms of maximum acceptable factory cycle time) depends on the variability imposed by factory starts, interarrival rates, and effective process times.

Now that the original, fully coupled version of this factory has been analyzed, we may proceed with an effort to reduce its DoR, that is, either fully or at least partially decouple the factory depicted in Figure 10.1. For sake of discussion, we shall partially and arbitrarily decouple this factory by means of adding three machines, one each to workstations A, B, and C, and then assign the machine-to-process-step dedications shown in Table 10.3. In this table, the three new machines are labeled as "Anew," "Bnew," and "Cnew." The decoupling has, as may be noted, resulted in three new virtual workstations: A′, B′, and C′.

The resulting workstation-centric flow plot for the reconfigured factory is shown in Figure 10.2. This partially decoupled factory has a DoR of 13/7, or 1.86. The nest, composed of workstations A′, B′, C′, and D, has a DoR of 10/4, or 2.5. The original factory had a DoR of 3.25, and thus the partial decoupling has resulted in a substantial reduction in the factory DoR (i.e., from 3.25 to 1.86).

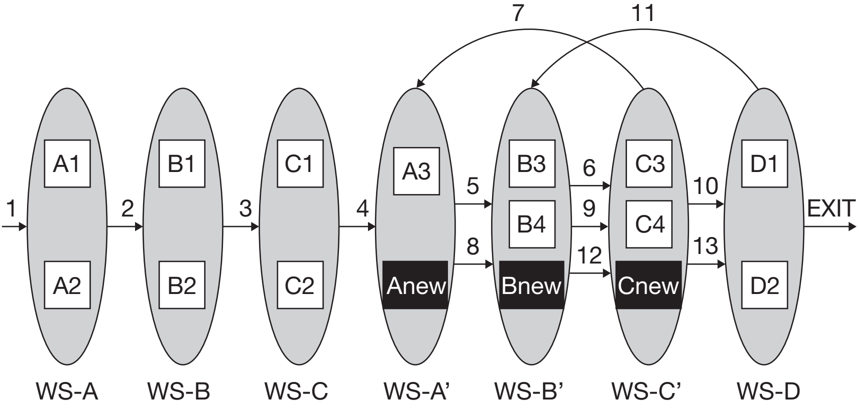
TABLE 10.3

Workstation Process-Step Dedications for Partial Decoupling

Workstation	Machines	Process Steps
A	A1, A2	1
B	B1, B2	2
C	C1, C2	3
A′	A3, Anew	4, 7
B′	B3, B4, Bnew	5, 8, 11
C′	C3, C4, Cnew	6, 9, 12
D	D1, D2	10, 13

FIGURE 10.2

Partially decoupled factory.



As accomplished previously for the original factory configuration (see Table 10.2), we may determine the composite arrival rates, maximum theoretical capacities, and occupation rates for the partially decoupled factory. The results are provided in Table 10.4.

So, having reduced the factory's DoR, have we improved its performance? The answer to this may be determined by either simulation or a pilot study. A simulation-generated comparison of the performance of the original versus the partially decoupled factory is shown in Table 10.5. (It should be noted that the availabilities and performance measures, as well as variability, were assumed identical in each factory configuration.)

TABLE 10.4

Workstation Maximum Theoretical Capacities EPR_{ws}

Workstation	Process Steps Composite Supported	Composite Arrival Rate (Lots/Hour)	Workstation Theoretical Capacity (Lots/Hour)	Workstation Occupation Rate
A	1	1.5	3.000000	0.500000
B	2	1.5	3.000000	0.500000
C	3	1.5	3.000000	0.500000
A′	4, 7	3.0	3.130000	0.958466
B′	5, 8, 11	4.5	4.629000	0.972132
C′	6, 9, 12	4.5	4.765000	0.944386
D	10, 13	3.0	3.272727	0.916590

TABLE 10.5

Comparison of Original and Partially Decoupled Factory

	Original Factory	Partially Decoupled Factory
Mean cycle time (hours)	30.050	23.480
Standard deviation of cycle time	6.270	4.400
Coefficient of variability of cycle time	0.210	0.189

By adding three machines to reduce factory DoR, factory cycle time has been reduced by 22 percent, whereas the variability in factory outs is reduced by about 10 percent. The question as to whether or not the reduction of DoR is worthwhile depends on a comparison of the cost of the additional machines versus the increase in profit (and customer satisfaction) resulting from the reduction in cycle time.

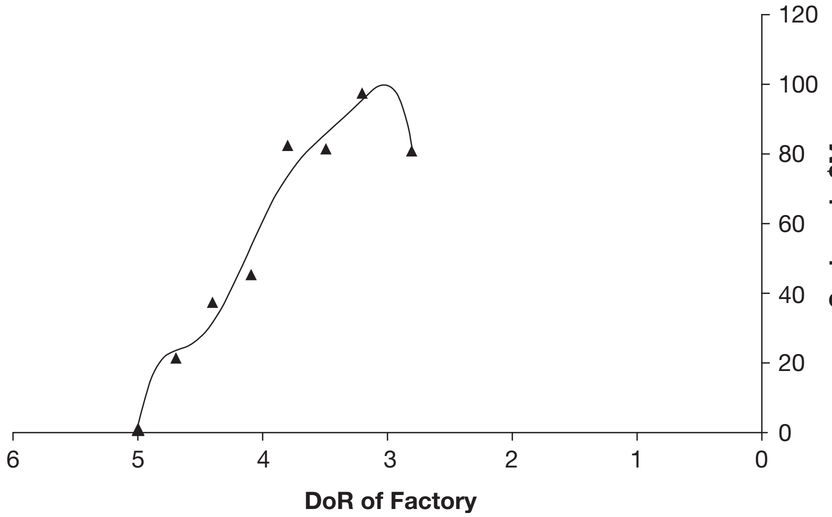
In a real situation, as opposed to this illustration, it may indeed be worthwhile to add machines to decrease cycle time, cycle-time variability, and customer lead time. For example, the cost imposed on a product over its lifetime by one additional day of cycle time in the semiconductor wafer fabrication industry is estimated (based on data in the open literature) to range from $5M to $100M depending on market conditions. Using the lower, more conservative value, a reduction in factory cycle time of, say, 10 days reduces cost by about $50M. If this offsets the cost of any additional machines, it would be worthwhile to reduce the DoR.

To more fully appreciate the potential benefit of DoR reduction, consider its implementation in an actual factory. This particular factory had an initial DoR of approximately 5.0. It was estimated that for every day of reduction in factory cycle time, the firm could—conservatively—increase its overall profit by $10M. Given data on the costs of any additional machines, an optimization procedure (Ignizio, 1992b) was employed to determine the optimal factory configuration for various budget constraints (i.e., constraints on the amount of money that could be spent on additional machines). The results of the optimization then were validated by means of factory simulations. The savings achieved via DoR reduction are shown in Figure 10.3.

This figure indicates that each reduction in factory DoR, from its original value of 5.0 down to 3.0, monotonically increases overall factory cost savings (i.e., considering both the cost of additional machines and the profit increase as a consequence of cycle-time reduction). Once, however, the DoR is reduced below about 3.0, the cost of the additional machines (i.e., those required to reduce the DoR to 3) begins to reduce the benefits from its maximum level of $100M. Clearly, it is unwise to ignore the potential benefits of a reduction in factory DoR.

FIGURE 10.3

Savings via process-step decoupling.



MAINTENANCE SPECIFICATIONS

The allocation of maintenance resources is one of the least appreciated and most underestimated factors in factory performance. I have labeled maintenance the "thankless science" because it receives so little respect and support in many firms. Yet improvements in maintenance protocols—and their supporting documentation—provide an enormous lever for the potential enhancement of factory performance (Duffuaa, Raouf, and Campbell, 1999; Robinson and Ginder, 1995).

Because of a failure to appreciate the importance of preventive maintenance (PM), maintenance costs often are substantially higher than need be, and maintenance resources are misdirected and overall factory performance suffers. Based on related surveys and personal experience (Anderson, no date; Ignizio, 1999), it is estimated that

- More than half the unscheduled downtime of machines is caused by errors in the most recent or previous PM activity.
- Because of poorly written PM specifications, time must be devoted to passdowns between factory shifts and to even the repeat of PM steps.
- A significant portion of unscheduled machine downtime is a result of a "hurry up" mentality (e.g., hurry a PM event so as to complete it before the end of a shift or before the lunch break).
- Calendar-based PM events (e.g., weekly, monthly, or quarterly) are employed mostly, if not solely, for convenience. If such events were scheduled less frequently (e.g., every nine days instead of weekly or every 35 days instead of monthly), most facilities could reduce maintenance costs by 10 to 20 percent—with no measurable impact on factory performance.
- Even in factories where maintenance activities are carried out relatively effectively, the personnel used to conduct the efforts are often misallocated (resulting in increased wait time as well as variability in downtime).
- Rarely is management aware of problems in maintenance protocols. Instead, managers tend to attribute poor machine

availability to physical sources (i.e., their focus is restricted to just the first two dimensions of manufacturing).

Anderson, in a Web-hosted paper summarizing a survey of this topic (Anderson, no date), notes that

- 30 to 40 percent of PM costs are spent on assets with negligible failure impact.
- A review of preventive maintenance activities indicated that
    - Only 13 percent of existing maintenance activities were worthwhile.
    - 19 percent of PM activities were a waste of time.
    - 30 percent of PM activities were carried out too frequently.
    - 70 percent of PM costs were incurred by just 25 percent of the PM activities.
    - Just under half of PM activities accounted for 90 percent of PM costs.

Such problems often are attributed to nothing more than abysmal (in terms of both content and language) PM specifications. One of the cheapest and most effective ways to improve factory performance is, in fact, to strive for C^4U-compliant maintenance specifications. Recall that C^4U-compliant specifications must be

- Clear
- Concise
- Complete
- Correct
- Unambiguous.

The trick, of course, is to translate these five adjectives into actions. This may be accomplished by the techniques employed by C. H. Waddington's group during World War II (Waddington, 1973) combined with lessons learned in later military and space programs (particularly the Apollo manned moon landing effort).

C^4U-compliant maintenance specifications may be achieved by following these steps and guidelines:

1. Cite the precise goal or goals the maintenance effort in question is intended to accomplish. (If you cannot cite the precise purpose of a maintenance specification, it is

unlikely that the resulting document will support the primary intent of preventive maintenance, for example, to avoid unscheduled downtime.[2])

2. Recognize that the developer of the maintenance specification must be thoroughly familiar with the machine or machines for which the specification is to be designed. (Too often maintenance is the thankless science in a factory, and thus it is assumed that anyone can develop a specification.)
3. Develop the initial specification, and then refine and validate it by means of a series of dry runs (i.e., carefully structured practice runs). Repeat these until the developer believes that no further improvement is possible.
4. Once the developer is satisfied with the specification, have someone other than the developer engage in additional dry runs. (Despite what the maintenance expert and specification developer may think, vital steps invariably will be omitted, and ambiguities will be present. These are best caught by a novice in his or her attempt to follow the steps of the specification.)
5. Use the previous two steps to eliminate unnecessary steps and avoid unsafe actions. Document every known deviation, ambiguity, and problem encountered, and revise the specification accordingly.
6. Repeat steps 3, 4, and 5 until the specification is deemed safe, credible, and effective—and the need for any passdown or explanation is, hopefully, eliminated.
7. After completing step 6 (and not until that step has been completed), take any actions necessary to ensure that the workspace is organized to support the final, refined specification.
8. Monitor the performance of each PM activity, and revise and refine the associated specifications (and subsequent workspace organization) whenever necessary. No matter what type of factory or what type of machine, continual improvement in PM specifications is a necessity.

---

[2] In discussions with factory personnel at more than two dozen factories, I found that the specific goal of a random sample of their PM specifications could *not* be cited more than 60 percent of the time. In many cases there was no agreement whatsoever on the purpose of a given specification. Much of the time, in fact, I was informed that a PM spec was conducted simply because "it had always been conducted."

There is one other matter that must be considered if a firm is to achieve C^{4}U-compliant PM specifications and benefit from them. Specifically, the firm must recognize the importance of the development of such specifications and adequately and publicly reward their developers. It must be made crystal clear that the development and continual refinement of PM specifications are a high priority and valued assignment.

Unfortunately, in too many cases the opposite impression is given to factory floor workers and engineers. In some firms, personnel actually may believe (a belief quite possibly based on experience) that they will be punished if they request a refinement to a PM specification. To provide just one example, I received an e-mail from a factory floor technician in which he cited a problem with a particular PM specification at his firm. He and his coworkers had discovered numerous (and obvious) errors in the specification 15 years ago and proposed changes to correct the deficiencies. The changes were ignored by superiors because any alterations in a specification had to be approved by some seven individuals in the factory and then sent to a virtual factory committee (consisting mostly of managers bent on reducing costs rather than technical experts) for final approval. The red tape, bureaucracy, and lack of interest in changes in documentation ultimately served to silence even the most dedicated and conscientious employee.

The preceding guidelines and steps have cited the actions required to produce a C^{4}U-compliant PM specification. There are also, however, certain guidelines and goals for determining if the effort in question actually has produced such a specification. Among these are

- The spec must be capable of being conducted successfully by other technicians, ideally without the need for input or a passdown.
- The spec must be shown, to the degree possible, not to induce unscheduled downs and must satisfy all safety, ergonomics, and human factor requirements.
- The spec must be shown to enable the right components to be examined, replaced, or repaired at the right time by the right people using the right tools as located in the right place and applied in the right order.
- The conduct of the spec must make a measurable and positive difference. (This is typically noted by improvements in availability and mitigation of the Waddington effect and

an improvement of the M-ratio.) If not, one must ask, "Why is this specification being performed?"

- A mechanism for anonymous input must have been provided.

What is always surprising to those who dismiss the importance of maintenance—particularly something so seemingly mundane as maintenance documentation—is the fact that a dedication to the development of C^4U-compliant PM specifications can and will provide significant, sustainable, and cost-effective improvements in factory performance.

OPERATING SPECIFICATIONS

Precisely the same points made with regard to the development of C^4U-compliant PM specifications hold true for the specifications that guide factory operations. This remains true even as more and more manual operations are automated (e.g., dispatching of jobs to machines, insertion of the jobs, and removal of the completed jobs).

One of the more overlooked areas for factory performance improvement is, in fact, that of the conversion of manual operations to automated methods. What is too often done is nothing more than the automation of existing manual operations. Unfortunately, if what is being accomplished manually is inefficient or even incorrect, all that automation accomplishes is the ability to do the wrong thing faster.

WORKSPACE ORGANIZATION (DECLUTTERING)

One of the most widely promoted tools of lean manufacturing is the process designated as *CANDO* (or *5S*). The purpose of this concept is to encourage cleanliness, order, and safety in the workspace. A further objective is to improve a workstation's ability to support the operations and maintenance activities in the factory. These are admirable intentions and definitely worthy of consideration by any firm.

While I encourage all companies to conduct workspace organization efforts, whether they are called CANDO, 5S, plain-old industrial engineering, or whatever, two critical factors are often overlooked. First of all, workspace organization is most definitely an art. One's ability to perform a useful workspace organization effort depends on one's motivation, training, and—in particular—experience. As such, it is vital to seek the guidance of those with experience and a successful track record.

While the conduct of a workspace organization effort—even by novices—may and likely will produce exciting before and after photographs, the impact on factory performance may not be nearly so exhilarating. Unless the effort serves to directly support the correct procedures to be used in the workstation's maintenance or operations, the results will be either less momentous than expected or even counterproductive. This unfortunate result may be avoided by acknowledging a second critical factor.

Specifically, it is important to make sure that a workspace organization effort is conducted *after* first completing the process of achieving C^4U-compliant specifications (either maintenance or operations, as appropriate). This was called out in the seventh step in the guidelines given in the preceding section.

The importance of performing the workspace organization effort *after* determining the existence of C^4U-compliant specifications—and the difficulty in selling such an idea—may be illustrated by means of a real-world example.

In one firm, a newly formed lean manufacturing team was eager to demonstrate its effectiveness to management. Team members identified one exceptionally cluttered and chaotic workplace (one that supported the maintenance operations for a particular workstation) and conducted a CANDO effort. Once this effort was completed, the workplace was as clean, tidy, and orderly as the most immaculate hospital operating room one would hope to find.

The before and after photos of the workspace were posted throughout the factory and presented to the factory manager. Duly impressed, she authorized support for the continuation of CANDO efforts.

One skeptic—let's call him Thomas—however, carefully examined the performance of the workstation both before and after the CANDO event. He discovered that the workstation's availability, cycle time, and departure-rate variability were as bad as before the cleanup effort. A plot of the Waddington effect showed, in fact, that the phenomenon still existed and was as pronounced as before. In short, the CANDO effort simply had made it easier to conduct a set of particularly dreadful PM events faster.

When Thomas's results were presented to the lean manufacturing team, they were met with open resentment. The lean team had photographic "proof" of the effectiveness of their effort—proof readily and eagerly accepted by the factory manager. Pictures, they reminded the chagrined Thomas, do not lie. Besides, as they informed him, they were too busy conducting CANDO events to bother with the thankless and time-consuming task of improving PM specifications.

The message in this example is most definitely not that CANDO is a bad idea. Rather, if one is to take full advantage of any workplace organization effort, it is wise to first achieve C^4U-compliant specifications—and then organize the workstation to support those specs.

WORKSTATION RUN RULES

The run rules [a.k.a. *dispatch rules* or *work-in-progress* (WIP) *management methods*] employed at a workstation can have a significant impact on the performance of both the workstation and the entire factory. This impact is even more pronounced (and complex) when the workstation supports multiple process steps and reentrancy.

Consider, for example, the implant machines (i.e., ion implantation) used in the semiconductor manufacturing industry. Such machines ionize dopant atoms, which then are isolated, accelerated, formed into a beam, and bombarded on the surface of a semiconductor wafer (Van Zant, 2000). Gases typically serve as the dopants, and these are usually fluorine-based. The typical implant machine is large and expensive.

The machines in the workstation that serve to conduct ion implantation characteristically support numerous implant operations, each of which may employ a different ionization source. Some operations, however, cannot be performed on the same machine owing to the incompatibility of the sources. In other instances, there may be a requirement to wait several minutes—or even hours—before one operation (using one type of source) can be safely followed by another (using another type of source).

Attempting to dispatch jobs to such machines manually is extremely complicated and confusing to human operators. Even the automation of such run rules may not (and typically does not) result in optimal or even near-optimal workstation performance.

The ideal solution would be to have each operation supported by a unique machine or set of machines (i.e., establishment of implant machines in a sequence identical to the sequence of implant operations). Such a fully decoupled implant process, however, would likely require the purchase of many additional machines and might be impractical because of both the cost and size of such machines.

A more practical alternative may be accomplished by partial decoupling via use of optimized operation-to-machine dedications. In most semiconductor firms, however, the operation-to-machine dedications of implant machines (as well as those of lithography machines, etc.) are accomplished heuristically. That is, the dedications are determined by some combination of expert judgment, experience, intuition, and luck.

Unfortunately for these firms, heuristic dedications are invariably (and substantially) inferior to optimized dedications. I will defer this discussion to Chapter 13 because optimized operation-to-machine (i.e., process-step-to-machine) dedications may be best accomplished by an extension of the mathematical model to be employed for the determination of workstation capacity.

"BROWN BESS" AND WADDINGTON ANALYSIS

I discovered more than a decade ago that many, if not most, of the concepts introduced in this chapter may be illustrated by means of a discussion of the evolution and refinement of the operation and maintenance protocols employed for the "Brown Bess," a 0.75-caliber musket of the eighteenth and nineteenth centuries (Antil, 2006). Rudyard Kipling does an admirable job of introducing this legendary weapon in a poem (a sonnet enhanced by a number of double entendres):

> *In the days of lace-ruffles, perukes, and brocade*
> *Brown Bess was a partner whom none could despise*
> *An out-spoken, flinty-lipped, brazen-faced jade,*
> *With a habit of looking men straight in the eyes*
> *At Blenheim and Ramillies, fops would confess*
> *They were pierced to the heart by the charms of Brown Bess.*
>
> —*Rudyard Kipling, 1911 "Brown Bess"*

Any reader who at this point may be wondering how on earth a discussion of an antique firearm could possibly add insight to the running of a factory must bear with me. I can only assure you that the Brown Bess discussion provides a concise and simple analogy that may be used to explain the purpose of a Waddington analysis (i.e., the procedure recommended for reducing complexity in operations and maintenance specifications and procedures).

First, however, I need to provide a brief overview of the Brown Bess musket—a "machine" considerably simpler than most readers would find on the factory floor, but one requiring very much the same type of support.

The Brown Bess was relatively reliable and cheap to manufacture and maintain and was not replaced until introduction of the percussion cap (in place of a flintlock) and rifles (i.e., a rifled barrel as opposed to the smooth bore and thus less accurate barrel of a musket). The musket, as a consequence of its poor accuracy and limited range, typically was employed by a force of musketeers arranged in carefully positioned ranks of men two to three deep. Firing of the musket, a black powder weapon, was conducted via a simultaneous volley of shots emanating from either all or just one row of men at a time.

Using such a system, and despite the weapon's inaccuracy (or any lack of sharpshooting skills by the musketeers), at least some of the enemy forces should be struck (i.e., "pierced to the heart by the charms of Brown Bess") in each volley. The primary components of such a musket are its wooden stock, smooth-bore metal barrel, rammer, and lock mechanism. The lock mechanism, the heart of the weapon, consists of a hammer (designed to hold a piece of flint), a priming pan (into which a priming charge of black powder was placed), and a frizzen (an L-shaped metal plate that, when closed, covered the priming pan and charge). Next to the priming, or flash, pan is a touch hole drilled through the barrel into the space where the main black powder charge is loaded.

When the trigger is pulled, the hammer snaps forward, causing the flint to scrape against the face of the frizzen, which, in turn, throws the frizzen back to expose the powder in the flash pan. The force of the flint on the metal produces a shower of sparks that are released into the pan, thus igniting the powder and sending flames through the touch hole. This results in ignition of the main charge of black powder in the barrel and fires the musket ball from the barrel.

Unlike the carefully planned, systematic, and focused conduct of a (properly designed and efficiently performed) Waddington analysis, enhancement of the operating and maintenance protocols of the Brown Bess evolved over decades through, in many cases, nothing more than trial and error or the inspiration of an individual musketeer. The operation of the musket is reflected in the steps required for its employment in battle. One early set of these steps (comprising the drill used in both training and battle) for a force consisting of two lines of musketeers (typical of the British) is as follows:

1. Order: "Prime and load" (accompanied by drumbeat). *Note:* In each step, the musket is held in the left hand while the right hand performs the operations.
2. Bring musket diagonally across your front, holding it midlength with the left hand (muzzle is high and to your left and does not point forward—a safety procedure).
3. Push the frizzen forward to open the priming pan (i.e., the flash pan).
4. Move the hammer to half cock.
5. Insert a charge of an "appropriate" amount from your powder horn into the musket priming pan.
6. Shut the pan with the frizzen so as to contain the priming powder in place.
7. Cast about, bringing musket diagonally across body in the opposite direction (muzzle is now high and to your right).
8. Use your powder horn to place a "sufficient" amount of gunpowder into the muzzle.
9. Remove a musket ball from its container (e.g., located in a bag carried in various locations on the soldier's person), and insert the ball into the barrel of the musket.
10. Stuff paper into the muzzle to serve as wadding (i.e., to keep the musket ball from rolling out of the barrel).
11. Seize the end of the rammer, and withdraw it from its storage location under the barrel of the gun.
12. Reverse the rammer, and ram down the ball on top of the charge (repeat this three times).
13. Withdraw the rammer and replace it in its storage location.
14. Bring musket to "Poise." It should be high on your left side, trigger facing out, held at neck by the left hand, with the right hand resting against it. Wait for next command. *Note:* With the musket at poise, the next order might be to "Shoulder," "Advance," or "Trail."
15. If the musket is to be fired, the next order is "Make ready," in which you bring the hammer to full cock using your right thumb.

16. If you are in the front rank, stand ready at "Poise." On the command "Make ready," kneel on your right knee, bring the hammer to full cock, and rest the musket butt on the ground near your knee.
17. Next order is "Present." Level your musket, and point it horizontally in front of you.
18. You may be ordered to "Oblique left" or "Oblique right," in which case all muskets point to either the left or the right.
19. At the order "Fire," pull trigger and then return the musket diagonally across your body, muzzle high and to the left; that is, return to step 2 and make ready for a repeat of the steps.

Now imagine, if you will, the training, practice, and nerve required to follow each of these steps in the face of a frontal attack by a column of determined French infantry, each shouting *"Vive l'emperor"* and accompanied by the rallying roll of their drummers. Add in the deafening roar of the muskets in the lines in front or behind you, coupled with the thick, acrid smoke of those black powder weapons, and you have a situation that definitely requires some precise and effective operating and maintenance protocols.

In addition, and vital to the effective conduct of the battle, was the cycle time of the musketeers, that is, the number of shots that could be fired per minute. During training exercises, a rate of fire of four shots per minute usually could be achieved. During an actual battle, however, the rate of fire might be as low as two shots per minute—unless shortcuts could be found to increase the effectiveness of the troops.

In short, the military was (or should have been) always alert for ways to reduce and/or refine the process steps employed in the drill. Stated another way, a reduction in the complexity of the drill (i.e., its process steps) was a primary goal—precisely as it is (or should be) in factory operating and maintenance "drills."

Over the years, the input of soldiers in the field and their officers and the inventions of weapons makers achieved a reduction in complexity and an increase in effectiveness. Some of this reduction was achieved by subtle physical changes; others by equally delicate changes in protocols. Just a few of these are as follows:

- In place of loose musket balls and separate strips of wadding, a paper (to be used later as wadding) twisted at the ends and enclosing a musket ball, the priming charge, and the powder to be inserted into the musket barrel was developed. The paper container itself was designated as a "cartridge."

- A container for the cartridges was designed and held in place on the right side of the musketeer (remember, the right hand was used for all operations and the left for holding the musket). This allowed for the replacement of powder horns with cartridge belts and, later, cartridge boxes (again, placed on the right side of the soldier).
- Instead of removing the musket ball and holding it in your hand, place it between your teeth (i.e., "bite the bullet"), and once powder has been inserted into the musket's barrel, spit the bullet into the barrel opening.[3]
- In actual battle and under extreme conditions, the soldier might ignore the loading and tamping of the wadding into the barrel. Instead, he would tap the butt of the musket against the ground to seat the bullet. (While this increased the rate of fire, it also reduced the velocity and range of the bullet.)

These refinements in the operating drill ultimately produced a reduction in process steps by about 20 percent and a subsequent increase in shots fired per minute.

Another change in protocols that may have achieved an even more substantial increase in the firing rate was accomplished in sixteenth-century Japan. Oda Nobunaga, a warlord credited with the introduction of firearms into Japan, used a unique and effective set of musket drill protocols.

This change in protocols turned Nobunaga's forces into particularly efficient killing machines. Nobunaga's favorite saying, by the way, was, "If a bird doesn't sing, kill it."

Nobunaga's musketeers worked in teams of loaders and shooters. Three muskets, plus the necessary powder, bullets, and supporting accoutrements, were assigned to each team. Immediately

---

[3] There are conflicting views as to whether or not the musketeer actually spat the bullet into the barrel. Some believe that this is a myth and that the soldier actually removed the bullet from between his teeth and manually inserted it into the barrel. Insertion was accomplished in this way, it is alleged, because after repeated firing, the musket barrel would be too hot to touch to one's lips. In lieu of either photographic evidence or a live (and exceptionally old) musketeer from that time period, I will leave it to readers to decide which view is correct.

after each shot, a cocked and loaded weapon was handed to the shooter by a loader. This protocol substantially increased the rate of fire of each musketeer.

While such a team effort was employed in Japan, for whatever reason, it never took hold in Europe or America.[4] Tradition, it would seem, can be a significant obstacle to change.

It should be noted that Frank Gilbreth developed an analogous team-based operating scheme in the early 1900s for protocols in hospital operating rooms. Gilbreth observed and took movies of actual operations. He noted that the "cycle time" of the surgeon was a major factor in the successful outcome of any operation. During steps of the operation, the surgeon would take his eyes off the patient and reach for whatever surgical instrument was needed for the next step. In fact, the surgeon would spend more time searching for instruments than in performing the operation. Gilbreth recommended that the surgeon focus on the operation and let a nurse hand him instruments. By nothing more than this simple change, the time required to complete an operation was reduced dramatically.[5]

The same team-based method, sometimes designated as the *nurse-surgeon-operation room* (NSOR) concept can and has been employed in factories and maintenance facilities. For example, an apprentice maintenance technician may serve as the "nurse," while the expert technician is the "surgeon." The expert focuses his or her attention on the maintenance event, while the apprentice reads off the steps of the specification and hands the expert the appropriate tool at the appropriate time.

When one factory agreed to implement a pilot study of the NSOR concept, the time to conduct PM events in several workstations was reduced by anywhere from 20 percent to nearly 50 percent. Equally important, the variability about these PM events was reduced by roughly two-thirds. Unfortunately for the

---

[4] One reason for the hesitancy to adopt such a team-based approach evidently was the belief that every man on the battlefield should be firing a weapon. This conclusion conveniently ignored the fact that much of the time the musketeer was engaged in reloading, repairing, or maintaining his gun. A modern-day team-based approach for the firing of a weapon is employed by sniper teams. One individual serves as the spotter (and also may estimate distance and wind direction and strength). Another is the shooter, the individual tasked with actually firing the sniper rifle at the target specified by the spotter.

[5] Like so much of Gilbreth's work, this concept has been rediscovered and renamed by those with little or no appreciation of the contributions of the pioneers of scientific management.

firm, the factory was unionized, and the effort was halted owing to the concerns of union leaders. They believed (possibly with good reason) that the reduction in time to conduct the PM events would result in a corresponding reduction in the maintenance workforce.

But let's return to the discussion of Brown Bess. The discourse until now has dealt with the reduction in the complexity of the musket drill (i.e., the operating steps). Equally important was the reduction in the complexity (and downtime variability) of the maintenance and repair of the musket, particularly in the heat of battle.

Both scheduled (i.e., more accurately, usage-based events) and unscheduled maintenance events were to be expected before, during, and following a firefight. These included such events as

- Loss or breakage of the flint
- Buildup of residue on the flint (with a subsequent reduction in the intensity of the sparks emitted)
- Blockage of the touch hole (a small hole that permitted the powder explosion in the flash pan to detonate the powder charge under the musket ball in the barrel of the musket)
- Jamming (i.e., the musket ball and wadding might become wedged partway down the musket barrel)
- Caking of residue in the musket barrel (after a number of shots, the residue left by the black powder would build up in the barrel and have to be removed)

The musketeer was, by necessity, both the operator and maintainer of his musket. Through training, observation, and word of mouth, a set of maintenance procedures was developed (e.g., the best-known methods for quickly and reliably replacing the flint, cleaning the flint, unblocking the touch hole, dislodging jammed musket balls, and removing residue in the barrel). There is, by the way, nothing like being shot at to encourage a speedy way to return one's weapon back to service.

It is important to recognize that the conduct of these maintenance events served to promote changes in the organization of the musketeer's "workspace." For example, over a period of time, the tools needed for maintenance and repair were identified, and their precise placement on the musketeer's person was determined. Each of these tools was placed in the soldier's cartridge box in precisely specified positions. The tools (and spare parts) necessary for maintenance consisted of

- A rag or brush to clean the flint and flash pan
- Spare flints
- A vent pick (to clean the touchhole)
- A screwdriver (to replace or adjust the flint)
- A bullet extractor (to remove jammed musket balls)

Over time, the bullet extractor evolved into a combination hammer, vent pick, and screwdriver (i.e., an early "all-in-one" tool set).

One other truly inspired approach to maintenance process-step reduction should be mentioned. As noted earlier, after prolonged firing, the barrel of the musket would become caked with powder residue. Furthermore, it should be pointed out that water boys were used to carry water to the dehydrated musketeers during breaks in the firing. As a result of drinking the water, the musketeers, quite naturally, found it necessary to relieve themselves.

A few ingenious musketeers found that they could both relieve themselves and remove the residue built up in the musket barrel. Simply and bluntly, they urinated into the barrel of the musket. By combining two "maintenance" steps into one, the downtime of the Brown Bess was reduced.

The reader should note that the story of the Brown Bess confirms the importance of improving a maintenance (or operating) specification *before* performing a workplace organization effort (e.g., CANDO). The organization of the workplace (in this case, the cartridge box and tools carried by the musketeer) was determined by and after changes in the operating and maintenance protocols employed rather than the other way around. While there is considerable resistance to this argument, history and successful improvements in factory performance prove its correctness.

At any rate, the reduction and refinement of operating and maintenance steps served to increase the firing rate of the musketeers substantially. The motivation behind these improvements was literally one of life and death. The motivation behind factory performance improvement may well be one of the life and death of the manufacturing firm. The faster the reduction in complexity of operating and maintenance procedures and specifications is accomplished, the quicker the factory will achieve (and exceed) world-class manufacturing status.

CHAPTER SUMMARY

This chapter discussed some of the more obvious sources of complexity within the factory. They are, however, by no means the only ones. There are, in fact, forms of complexity even more widespread and sometimes more damaging. Among these are

- Complexity in the factory's and firm's business processes
- Complexity in the firm's supply-chain network and protocols
- Complexity in terms of the inconsistency and ambiguity in the firm's mission statements, goals, requests, and pronouncements
- Complexity induced by frequent changes in the firm's goals, measures of performance, and mission
- Complexity in terms of the retrieval of data and information
- Complexity in the number of steps and red tape required to make a change in maintenance and operation protocols

In other words, any team involved in complexity reduction (whether labeled a lean team, industrial engineers, factory performance team, a *kaizen* group, etc.) should look beyond the more obvious factory protocols and flow paths. Furthermore, the same concepts that may be employed to reduce complexity in these areas may be extended to virtually all other facets of the firm, and vice versa.

CASE STUDY 10: MIDCOURSE CORRECTION

Ben and Donna take a seat and wait, apprehensively, for Tommy to announce the reason for the meeting. Tommy, for his part, simply sits there, glaring at them. Donna decides to break the uncomfortable silence.

"I couldn't convince that stupid professor to tell us anything about simulating by means of fluid models. But I did have one of my people look into the matter. He says that it's pretty straightforward, but it would take a long, long time to construct such a model for this factory. I figure we can convince Winston Smith to do the grunt work on that matter. But I'm not quite sure just what we do with the model once it's been built."

Tommy continues to glare.

"Folks," says Ben, "all we need to do is to threaten Winston once again. Just tell him that if he doesn't explain how building this fluid model will decrease our cycle time, we'll get rid of Julia."

"You two idiots," Tommy replies, "I've already checked with our director of manufacturing, and he says that the introduction of a simulation package other than what has been authorized would be a violation of the 'No Deviations' policy. Besides, this firm is no longer concerned with reducing cycle time. What Marvin Muddle wants us to do now is to increase our capacity, and to hell with cycle time. It seems that every factory in our system has reduced its factory starts so as to reduce cycle time. We all may be fast now, but we're not able to fulfill our customers' orders. Every Muddle factory has been ordered to increase factory starts. Cycle time is now a moot point. So now, you two geniuses, our mission is to increase factory capacity. Ben, send out an e-mail to our factory managers. Tell them we have to increase our starts to 11,000 units per week. And you, Donna, forget about that idiot professor and this nonsense about fluid models. I want you to increase workstation availability and utilization by 20 percent. This meeting is over."

* * *

A day later we find Julia, Winston, Dan, and even Brad in the "war room." Julia has some news.

"Fellows, Tommy Jenkins has ordered his department managers to forget about factory cycle time. The mission now is to increase capacity to 11,000 units per week. He also has demanded an increase in workstation availability and utilization. I guess we shouldn't be surprised. Marvin Muddle has a short attention span. It's cycle time one day, capacity the next, then moves, and then Marvin Muddle's all-time favorite, cost reduction."

Winston rolls his eyes and replies, "The maximum sustainable capacity of this factory, under near-perfect conditions, is about 10,000 units per week. They may start whatever number they want, but the weekly output will be no more than 10,000 units. Besides, within a few weeks, there will be no room for the inventory that will be built up in the queues in front of the factory workstations. This will just be one more disaster. Don't they ever learn?"

"Apparently not," Julia replies. "But is there any rational way to increase our factory's capacity? Can we use your fluid models to investigate that?"

"Certainly," says Winston. "There are a number of ways to increase capacity, some better than others. As you may have learned in your lessons with Aristotle, there is a cost-effective way to improve factory performance. Just reduce complexity and variability. Do that, and you're guaranteed to improve overall factory performance, including both cycle time and capacity."

"You've shown that the declustering of factory starts will reduce cycle time. I noticed that it also results in a slight increase in capacity. Is there something else that could improve capacity even more?" asks Dan.

"Yes," Winston replies. "One almost sure-fire way to increase workstation availability is to employ a Waddington analysis. In particular, you want to have PM and operating specifications that are C^4U-compliant. The only problem is that you can't demonstrate this change with any type of simulation model. Another thing you can do is to optimally allocate maintenance personnel. These things, and others, will increase availability. If you simultaneously reduce complexity and increase availability, you can get some truly substantial increases in factory capacity. Hasn't Aristotle discussed these approaches with you?"

"Yes, he has," says Dan, somewhat sheepishly. "I have to admit that I had forgotten about it. But how do we get an approval to conduct a Waddington analysis?"

"I can answer that," says Julia. "It isn't going to happen. There's no way we'd be able to convince management that we should assign people to improve our maintenance and operating specs or to even consider the reallocation of our maintenance technicians. These are all serious violations of the 'No Deviations' policy. Besides that, the quality control group would go wild. It's not going to happen."

"Julia's right," says Brad. "The only way we could get an approval for any of this is to introduce an educational effort—have the professor and Winston present the material on the science of manufacturing. Maybe then they'd finally listen."

"That's not going to happen either, Brad," Julia replies. "Management is not going to permit courses on the science of anything. We'd have to condense everything into a slogan and maybe 7 to 10 principles. And those would have to be made as simplistic as possible."

"So," says Dan, "what do we do? Nothing?"

"Let's talk to Aristotle," says Winston. "Maybe he has some ideas. In the meantime, did any of you notice that the cubicle that was put in a few days ago, along with its occupant, has suddenly vanished?”

CHAPTER 10 EXERCISES

1. Given the four-workstation factory in Figure 10.1 and the data in Table 10.1, develop a *fully* decoupled factory capable of handling the 1.5 jobs per hour arrival rate.
2. Compute the occupancy rates of each workstation in the fully decoupled factory of Exercise 1.
3. List at least three actual examples of accidents that resulted (or likely resulted) from a failure to develop C^4U-compliant PM specifications. (*Hint:* Look into accidents in the space program, airlines, automobiles, and construction.)
4. Perform the following experiment: First, ask a colleague to add paper to a copy machine. Then record, in detail, the steps followed. Next, determine if there is a more efficient way (e.g., by reducing the number of steps).

This page intentionally left blank

CHAPTER 11

Reducing Variability

While complexity indirectly (and sometimes substantially) induces variability, there are other more direct (i.e., first- and second-order) sources of this particular enemy of factory performance. As noted in Chapter 5 and exhibited in the 12-workstation demonstrations of Chapters 4 and 6, variability increases cycle time, increases factory inventory, and subsequently serves to reduce the factory's maximum sustainable capacity. The three sources of variability explicitly cited (i.e., first-order sources) in the second and third fundamental equations of manufacturing are

- Variability in arrival times (i.e., interarrival rates) of jobs entering into the queue in front of a process step and designated as C_{AR}
- Variability in departure times (i.e., interdeparture rates) of jobs exiting a process step and designated as C_{DR}
- Variability of the process-step effective process times and designated as C_{EPT}

These three principal sources of variability are themselves determined by a variety of second-order causes, including but not limited to batch sizes, factory starts protocols, macro and micro work-in-progress (WIP) management protocols, operating protocols, maintenance protocols, and spare parts and supplies protocols.[1] The most practical and cost-effective means to reduce factory

---

[1] For example, the variability about effective process times C_{EPT} is determined in part by a second-order source of variability, the variability about repair and maintenance events, that is, C_{DE}.

variability must address these second-order causes and are the subject of the sections that follow.

BATCH SIZES AND VARIABILITY

Batching serves to increase both arrival- and departure-rate variability. Recall the example in Chapter 5 in which batching (with a relatively small batch size of four) at a previous process step increased arrival-rate variability (i.e., C_{AR}) at the subsequent process step by more than three times what it had been with no batching. The larger the batch size used by the machines supporting a given process step, the greater is the arrival-rate variability at the next process step.

As a consequence, a particularly effective way to reduce variability is simply to reduce batch sizes throughout a factory. If, however, a process step (e.g., such as heat treating) takes a long time, say, eight hours, to complete, the temptation (and "quick and easy solution") is to use batching. For example, the heat-treatment furnaces employed in semiconductor manufacturing may use batches of wafer lots of sizes four, five, or more. Furthermore, each individual lot of semiconductor wafers may contain a dozen or more wafers (lot sizes of 25 wafers are not uncommon).

Thus, if a batch size of six lots (of 25 wafers per lot) is employed, the total time to process the 150 wafers would be eight hours. However, if the same furnace were used to process just a single wafer at a time, the 150 wafers would require 1,200 hours (i.e., 50 days) in total—just for this one heat-treatment step.

Consider, however, what is at this time the Holy Grail of semiconductor manufacturing: single-wafer processing (SWP) (Ignizio, 2004; Wood, 1995). As noted earlier, if we had to rely on conventional furnaces, SWP would not be practical. Advances in furnaces using rapid thermal processing (RTP), though, may be used to make SWP not just feasible but advantageous.

Consider what a strictly hypothetical reduction in the process-step time of the furnace from 8 hours to, say, 10 minutes would accomplish. The six-lot batch (of 25 wafers each) discussed previously could be heat treated, one wafer at a time, in 25 hours in total if RTP were employed. This, of course, is still more than the 8 hours required with batching. However, since a RTP furnace is of smaller size (and, hopefully, lesser cost), we might employ 10 (or more) RTP furnaces for every conventional one. If this were done, the total process time of the 150 wafers using 10 RTP furnaces would be just 2.5 hours—a significant reduction over the 8 hours required for this process step using a conventional batching furnace. Quite possibly more important, however, is the reduction in variability achieved by SWP over batch processing.

The message is that whatever can be done to reduce (or, better yet, eliminate) batching should be considered. SWP coupled with RTP plus other advances offer one hope for the reduction of overall process time and variability via the elimination of batching—or at least a reduction in batch sizes. Another, even simpler way to improve factory performance is to reduce the variability imposed by inferior (and, unfortunately, rather typical) factory starts policies.

FACTORY STARTS PROTOCOLS AND VARIABILITY

One of the easiest and quickest ways to reduce variability is simply to decluster factory starts. A declustered factory starts protocol is one in which jobs (i.e., in either single units or, if necessary, lots or batches) are introduced into the factory in such a way as to minimize the coefficient of variability (CoV) of the interarrival rate of jobs into the queue in front of the very first workstation. This may be accomplished if jobs are introduced in a *smoothed* manner.

For example, if the desired factory output is 480 jobs per a 24-hour day, then the ideal time between the insertion of one job and the next would be 3 minutes. That is, we divide the number of jobs to be started each day (480) by the number of hours in the workday (24) and achieve a result of 20 jobs per hour or—using a smoothed starts protocol—one job every 3 minutes.

Few factories I have encountered, however, employ a smoothed starts protocol. Instead, mainly for reasons of alleged convenience, jobs are clustered. For example, in the preceding illustration, the goal was 480 jobs to be started per day. Some factories simply may insert 480 jobs in front of the first process step at the beginning of the day. Others might make a feeble attempt to decluster by, for example, introducing 160 jobs every 8 hours. While the first scheme will impose considerable interarrival-rate variability, the second (i.e., 160 jobs every 8 hours) is likely to be only marginally better.

Firms that have taken the time to decluster factory starts have seen substantial reductions in factory cycle time (particularly in the front end of the factory). Furthermore, if a declustered factory starts protocol is combined with an effective factory starts *load-management scheme,* reductions in total factory time will be even more pronounced.

A factory load-management scheme, in turn, is a protocol for reducing or increasing (i.e., synchronizing) the jobs started into the factory in accordance with the health of the factory (e.g., according to reasonably accurate predictions of the factory's constantly changing sustainable capacity). One of the worst mistakes made by a firm is to blindly follow some *a priori* quota for starts while ignoring the fact that the environment within any real-world factory is in a constant state of flux.

For example, if the quota established (by whomever, too often by the dictates of the finance department) happens to be 5,000 jobs per week, but the factory is only capable of handling 4,500 jobs this particular week (perhaps owing to the need to conduct preventive maintenance or repair on a bottleneck workstation), all we would achieve by unbendingly adhering to the quota is to diminish factory performance. Inventory and queues will increase, variability will increase, and factory velocity will decrease. And even if the factory's real capacity is returned to something greater than 5,000 jobs per week, it could take weeks to work off the inventory and queues built up in that single week of factory overstarts.

A word of warning: One of the potential reactions to any proposal for the smoothing of factory starts—or of synchronizing the imposed factory load in accordance with actual factory capacity—is to be told, "This will require too much time and thought." (One might be tempted to respond that if the goal is to reduce time and thought, why not just shut down the factory? Hopefully, however, any urge to voice such an imprudent reply will be resisted.) The usual reason a smoothed factory starts protocol is opposed is because such a change requires more oversight and planning than that of simply dumping the jobs to be started for the day in front of the first workstation at the beginning of each day. This resistance sometimes may be overcome by conducting either a simulation or a carefully structured pilot study (e.g., one that employs a smoothed starts policy for a month or more, followed by a comparison of the results with the original, clustered protocol).

Firms that have been convinced of the need for a smooth starts policy, coupled with a rational factory loading scheme, have seen reductions in factory cycle time by as much as 50 percent. When combined with the next topic, the smoothing of preventive maintenance (PM) events, reductions of as much as 70 percent sometimes have been achieved.

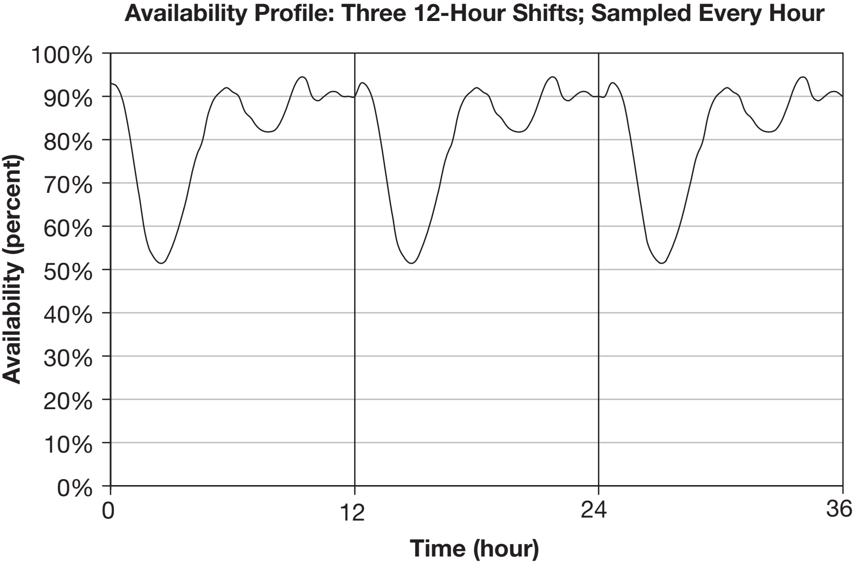
MAINTENANCE EVENT SCHEDULING AND VARIABILITY

Declustering of factory starts, as discussed above, is a quick, easy, and effective way to reduce variability. The same holds true for the declustering of PM activities. A declustered factory maintenance protocol is one in which maintenance events are scheduled in such a way as to minimize any induced variability and increase effective workstation availability. This may be accomplished if PM events are scheduled in such a way as to evenly spread out their resulting machine and workstation downtimes.[2]

In Chapter 8, a factory performance metric designated as the *availability profile plot* was introduced and illustrated by means of two graphs, repeated here as Figures 11.1 and 11.2. As may be seen, there are sharp decreases in workstation availability at the beginning of each 12-hour shift.

FIGURE 11.1

Hypothetical workstation availability profile plot.

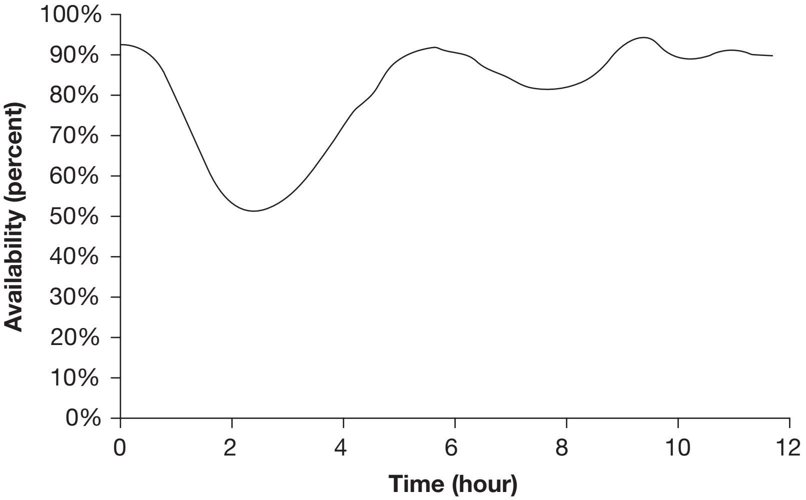


---

[2] A formal optimization method for the achievement of smoothed PM event scheduling (i.e., a method for the establishment of a PM schedule that simultaneously minimizes any overlap of PM events at a workstation *and* evenly spreads out these events) has been developed (Ignizio, 1978, 1992a, 1999).

FIGURE 11.2

Workstation availability profile plot, one shift.



As discussed in Chapter 8, a possible reason (one that actually has been identified in several firms) for a sharp dip in a workstation's availability at the beginning of a shift is the desire on the part of the workforce assigned to that station to avoid having a PM event extend into the next shift. Perhaps in the past workers in the next shift have complained to management about the number of incomplete PM events being passed on to them. Perhaps there is a desire on the part of the workforce on every shift to avoid lengthy pass-down meetings between shifts (e.g., to list the PM events that have yet to be completed and discuss their status). Perhaps the workers want as many PM events as possible completed before a lunch break. Or perhaps the PM specifications are so poorly written (i.e., they are not C^4U-compliant) that the maintenance technicians on the next shift are not confident that the PM steps presumably completed by the preceding shift have been conducted (or documented) properly. Any or all of these reasons serve to motivate the clustering of PM events at the beginning of a shift.

Time- or usage-based PM events usually are conducted during some window of time rather than at a fixed time. For example, a time-based PM event (say, event *X* on workstation A) may be scheduled for every 40 hours, but the actual window for the conduct of event *X* might be anywhere between 35 to 45 hours since the last time event *X* was conducted on a given machine. Given that workstation A consists of several machines, each having been in operation a certain number of hours since the last PM event *X*, the goal should be to evenly spread out the performance of these events while still staying within their desired window of conduct.

To clarify the concept of a smoothed (i.e., truly declustered) versus a clustered schedule of PM events, consider an example. To keep matters simple, assume that we have a workstation consisting of five machines. During the forthcoming eight-hour shift, each of the machines must undergo some type of PM event. Furthermore, assume that all these PM events must be initiated within the first five hours of the shift.

Table 11.1 lists the predicted time (i.e., ignoring variability and assuming ideal conditions) required to conduct each of the PM events. Note that the specific type of PM event may differ during the time period of interest for each machine (e.g., the PM required for machine 1 is different—in this shift—from that required for machine 2 and, for these cases, consumes 1.00 and 1.50 hours, respectively).

There are an infinite number of possible schedules for these five events, but Figure 11.3 presents the three most pertinent to this discussion. In Figure 11.3*a*, all five events are started at time zero (i.e., at the start of the shift), and as a consequence, the events are both overlapping and clustered. Such a schedule might be motivated simply by the desire to get the PM events conducted as soon as possible in the shift. Of the three schedules to be discussed, this is the worst possible choice. Unfortunately, for some firms, it is the first and only choice.

In Figure 11.3*b*, the PM events of the second approach have been scheduled to minimize (or, if possible, to eliminate) any overlap. The motivation for such a schedule likely would be to complete all maintenance events as soon as possible while maintaining as many machines in operation at any one time as feasible.

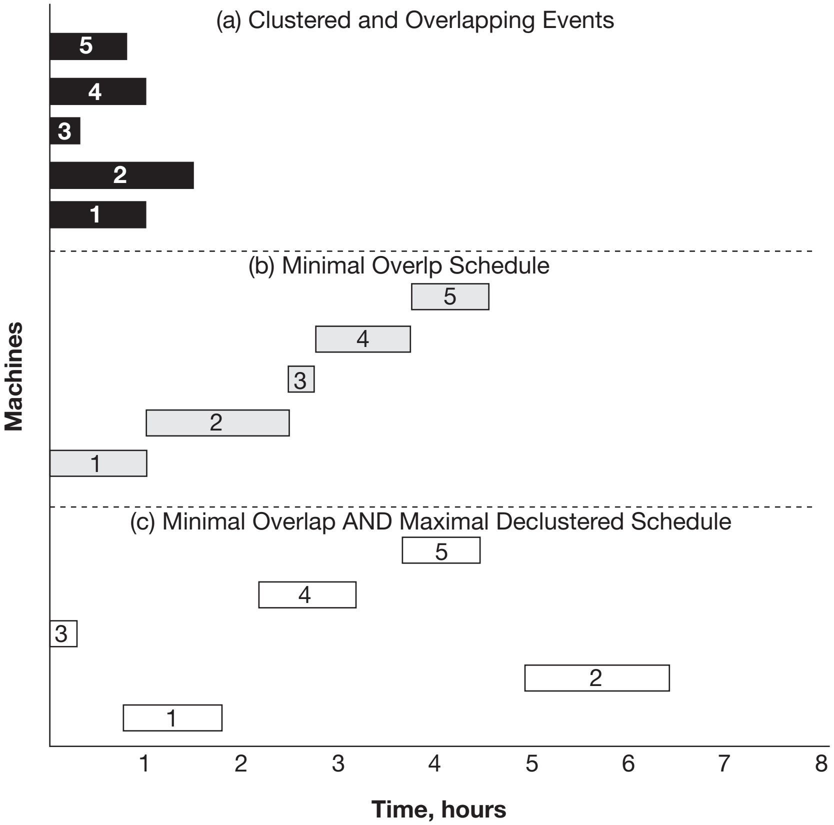
TABLE 11.1

PM Event Times per Machine

PM Event	Time Required (Hours)
Machine 1 PM event	1.00
Machine 2 PM event	1.50
Machine 3 PM event	0.30
Machine 4 PM event	1.00
Machine 5 PM event	0.80

FIGURE 11.3

Clustered, minimized overlap, and declustered PM event schedules.



Finally, in Figure 11.3*c,* a truly declustered schedule has been established. The difference between simply minimizing overlap and truly declustering PM events is apparent when the schedule of Figure 11.3*b* is compared with that of Figure 11.3*c.* While the schedule in Figure 11.3*b* has no overlap, it is still clustered; that is, the PM events take place contiguously between 0 and 4.6 hours, followed by 2.4 hours in which no events occur. Consequently, the PM events of Figure 11.3*b* are clustered, although not as severely as in Figure 11.3*a,* within the initial portion of the shift.

Now consider the impact on workstation availability of each of the three schedules shown in Figure 11.3. Whatever the schedule employed, the predicted workstation availability is 88.5 percent. Consequently, the factory manager or factory engineers (particularly if their primary focus is on average workstation availability) may be lead to believe that it doesn't make any difference as to the schedule employed. Thus, under the illusion that the maximally clustered schedule of Figure 11.3*a,* the minimal overlap schedule of Figure 11.3*b,* or the truly declustered schedule of Figure 11.3*c* result in identical workstation availabilities, they may see no reason to consider declustering the PM events.

If so, they are ignoring two significant factors. First, the PM event times in Table 11.1 are predictions—based on ideal conditions. However, unless there are a sufficiently large number of maintenance technicians (MTs), the actual PM times that will be experienced by either the maximally clustered or minimal overlap schedules typically will exceed those that are predicted. For example, in the maximally clustered schedule of Figure 11.3*a,* there must be enough MTs to simultaneously start and conduct the five PM events on the five different machines. This likely would require the firm to employ far more MTs than necessary.

A somewhat similar argument holds for the minimal overlap schedule in Figure 11.3*b.* While this schedule might require fewer MTs, it assumes that they can transit from a PM event on one machine to one on another in zero time. This is obviously unrealistic.

The second significant factor that serves to present problems when either a maximally clustered schedule or a minimal overlap schedule is employed is that of the likely buildup of a queue in front of the workstation because of such schedules. This will induce additional variability in the departures of jobs exiting the workstation. (This particular impact may be best illustrated by means of simulation.)

In summary, an improvement in factory performance may be achieved by declustering PM events. While one should expect some resistance to this "radical" notion, it is important to gain management and workforce support for declustering. The benefits of declustering overwhelm any real or imagined obstacles.

MAINTENANCE PERSONNEL ALLOCATION

The allocation of maintenance personnel to workstations plays a particularly significant role in factory performance. Inferior allocation schemes increase the time wasted in waiting for an MT or team of MTs to arrive and conduct a PM event or repair. This both decreases workstation availability and increases factory variability. The guidelines that should be employed, or at least seriously considered, with regard to the allocation of MTs to workstations include the following:

- Care should be taken to train the MTs properly and to equip them most effectively and efficiently to perform their duties.
- A factory cannot afford to tolerate the employment of any MT who either cannot or will not exhibit the skills necessary to perform his or her duties effectively and efficiently. Oversight and firm actions in support of this goal must be established.
- Consideration should be given, wherever and whenever possible, for the cross-training of MTs. Cross-training can play a significant role in reducing wait-for-MT times and variability.
- Avoid the allocation of MTs according to the "squeaky wheel syndrome." The "owners" or advocates of a given workstation may well be louder and/or more eloquent in their demands for MTs, but emotions and marketing skills should not be the basis for MT allocation. The focus always must be on the improvement of overall factory performance.
- An increase in factory capacity might be (and likely will be) accomplished by the allocation of additional MTs to constraint workstations (or, more properly, to the machines supporting the constraint process steps).
- The workstations receiving the highest priority for the allocation of MTs, however, should be the ones whose downtimes have the greatest impact on overall factory performance.

The last point in this list may be best achieved by means of an optimization model (Ignizio, 2004). Alternately, a reasonably effective MT-to-workstation allocation may be accomplished by adhering to a set of heuristic rules.

To illustrate, a hypothetical factory configuration—one in which each machine in a given workstation has performance identical to any other machine in that workstation—is employed. Assuming the existence of a valid factory simulation model—one that considers the allocation of MTs to workstations—the basic steps to be followed by the heuristic approach for this type of factory are as follows[3]:

---

[3] We shall let *CAP* represent the maximum theoretical capacity of the entity, that is, its *EPR* value. Later, once a more precise estimate for maximum sustainable capacity (*SC*) is developed in Chapter 13, then *CAP* may be replaced by the *SC* of the entity.

- Determine the value of the following parameters:
  - $m(ws)$ = number of machines in each workstation.
  - $CAP_m(ws)$ = capacity of each machine in the given workstation.
  - $CAP(ws)$ = capacity of all the machines in the given workstation (i.e., the sum, in this case, of the individual capacities).
- Let the bottleneck workstation's capacity be designated $CAP(bn)$, where bn denotes a bottleneck (i.e., a constraint).
- $CAP(ws,-1)$ = capacity of the workstation if one machine goes down.
- Determine the first weighting factor, designated $w(ws,1)$, for each workstation, where
  - If the workstation is a bottleneck, set its first weighting factor to a value of 1; that is, $w(bn,1) = 1$.
  - If the workstation is not a bottleneck, determine the value of x as follows:
    - $x = CAP(ws,-1) - CAP(bn)$
    - If $x \geq 0$, then $w(ws,1) = 1$.
    - If $x < 0$, then $w(ws,1) = \dfrac{CAP(ws)}{CAP(ws) + x}$.
- Determine the second weighting factor, designated $w(ws,2)$, for each workstation, where m is the number of machines in the workstation of interest:
  - $w(ws,2) = \dfrac{m}{m-1}$ if $m > 1$; otherwise, $w(ws,2) = 1$
- Determine the third weighting factor, designated $w(ws,3)$, for each workstation, where NO is the number of operations performed by the workstation of interest:
  - $w(ws,3) = \sqrt{NO}$
- Determine the fourth weighting factor, designated $w(ws,4)$, for each workstation as follows:
  - If the workstation is a bottleneck, then $w(ws,4) = 3$.
  - If the workstation is not a bottleneck, then $w(ws,4) = 1$.
- Determine the fifth weighting factor, designated $w(ws,5)$, for each workstation as follows:
  - $w(ws,5) = 1.5$ if the workstation directly feeds a bottleneck.
  - $w(ws,5) = 1$ otherwise.

- The composite weighting factor, designated $W(ws)$, assigned to workstation ws is given by
  - $W(ws) = \prod_{s=1}^{5} w\,(ws, s)$
- Determine (or estimate) the expected arrival rate (designated λ) of machine downs per unit of time for each workstation (e.g., if a machine fails every 50 hours, the arrival rate is 0.02 machines per hour).
- Determine (or estimate) the expected repair or maintenance rate (designated μ) per machine for each workstation (e.g., if it takes on average four hours to repair a machine, the rate is 0.25 machines per hour).

To illustrate, consider a workstation, say, workstation, with four machines and located within a factory employing a 168-hour workweek, where

- Each machine is capable of processing 500 widgets per week (i.e., $EPR_m = 500$).
- The number of operations supported by this workstation is nine.
- The workstation is not a bottleneck and does not directly feed a bottleneck.
- The capacity (*EPR*) of the bottleneck workstation is 1,900 widgets per week.

Thus, for this workstation, the pertinent parameters are

$$CAP(3) = 4 \bullet 500 = 2{,}000$$
$$x = (2{,}000 - 500) - 1{,}900 = -400$$
$$w(3,1) = \frac{2{,}000}{2{,}000 - 400} = 1.250$$
$$w(3,2) = 4/(4-1) = 1.333$$
$$w(3,3) = \sqrt{9} = 3.00$$
$$w(3,4) = 1$$
$$w(3,5) = 1$$

Consequently, the composite weighting factor is

$$W(3) = 1.25 \bullet 1.333 \bullet 3 \bullet 1 \bullet 1 = 5$$

The same process may be used to compute the composite weighting factor for all the workstations in the factory. Once this is done, you may use the λ and μ values of each workstation to compute expected wait times for the MTs. Assume that $\lambda = 1.2116$ and $\mu = 23.02$ (i.e., each machine in the workstation goes down at an average rate of one every 138.66 hours, and each maintenance event has an expected duration of 7.3 hours per machine).[4] Using the queuing theory, the expected time in hours spent in the maintenance queue for the workstation may be computed (Taha, 2006; Hillier and Lieberman, 2005).

Once the preceding analysis has been performed for each workstation, then you can conduct a series of factory simulations. Whenever factory performance is less than desired, reallocate the MTs—from the workstations having the lowest priority to those having the highest priority (with priority established by the values of the workstation composite weighting factors combined with the results of the supporting queuing analysis).

One way in which to determine this priority is to rank the workstations according to their composite weighting factors—and (as an option) factor in the queuing theory results. For example, a workstation with both the highest composite weighting factor and the most wait time for maintenance might be assigned the highest priority for a reallocation of MTs (e.g., move an MT from the lowest-priority workstation to the highest-priority workstation). Continue this procedure until no further improvement in factory performance appears feasible.

To demonstrate the importance of the proper allocation of MTs to workstations, consider the difference between optimal MT allocation and the traditional MT assignment policy employed by a firm. The total number of MTs, as deduced by the traditional method, was used as a baseline for comparison. These MTs also were allocated in the traditional manner (i.e., hunches, guesses, averages, and the "squeaky wheel syndrome"). This number was lowered gradually, and the cycle time for each reduction (i.e., in terms of the percent of original number of MTs) was computed by means of factory simulations.

Using the same total number of MTs, an optimization model (Ignizio, 2004) was employed for the MT-to-workstation

---

[4] These values are determined by (1) dividing 168 hours by λ and (2) dividing 168 hours by μ.

assignments. The total number of MTs then was reduced sequentially and their allocation reoptimized. Figure 11.4 summarizes the results.

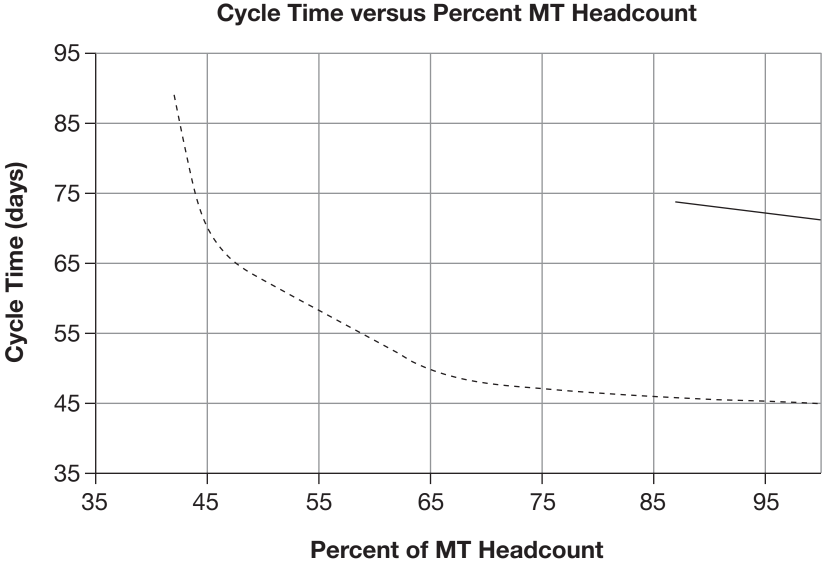
The figure shows as a solid line the cycle time of the factory using the traditional method of MT-to-workstation allocation. At 100 percent of the number of MTs (as computed or conjectured via traditional means), the factory cycle time is 71 days. If headcount is reduced to 87 percent of the original number, the factory cycle time is 74 days (but the factory becomes highly unstable). For any further reduction, the factory "breaks"; that is, cycle time goes ballistic.

On the other hand, by using the optimal allocation of MTs to workstations, the cycle time for the factory at 100 percent of the original number of MTs (again, as computed via the traditional method) is just 45 days. In fact, the total number of MTs (as computed via the traditional allocation) may be reduced by about 45 percent, and the factory's cycle time will be the same as that of the traditional method at 100 percent of MT headcount.

When the optimization method for MT-to-workstation allocation has been employed, factories have seen significant reductions in cycle times—anywhere from a 20 to 50 percent reduction—over

FIGURE 11.4

Comparison of cycle times.



that experienced under their previous heuristic approaches. The optimization method offers not just substantial factory performance improvement but also the opportunity to reduce factory headcount.

SPARES AND SUPPLIES LOCATION AND VARIABILITY

Just as an inefficient number and allocation of MTs to workstations degrades factory performance, so will an inappropriate number and location of spares and supplies. The case study in Chapter 1 (recall Dan Ryan's frantic attempts to obtain spare parts for his workstation) provided a hypothetical illustration of the problems a factory will face if it has too few locations from which to dispatch spare parts and supplies. While that story is fiction, the consequences of an inadequate number of dispatch stations are definitely real. The same consequences hold true if the dispatch stations are not optimally or at least nearly optimally located.

Specifically, inadequate numbers and/or inferior locations of dispatch stations increase the duration and variability of the wait times inflicted on workstations needing parts and supplies. Unfortunately, as in the case study in Chapter 1, some firms may believe that a reduction in the number of dispatch stations will save money—or that the locations of these stations do not play a significant role in factory performance. Such reasoning, however, is wrong on both counts.

A solution to determining locations and number of dispatch stations may be found by means of a well-designed simulation effort (assuming that the impact of dispatch station location is credibly portrayed in the simulation). A faster and more accurate means for deciding these matters is available through either optimization or heuristic models specifically designed for this problem. Adoption of such methods usually will result in significant and sustainable improvement in factory performance and also will indicate the impact of either reducing or increasing the number of dispatch stations. Certain heuristic approaches in particular are fast and simple to employ (e.g., the only mathematical operations in one are addition, subtraction, and comparison). More details on these methods, as well as numerical illustrations of their implementation, may be found in the references (Francis and White, 1974; Ignizio, 1971; Ignizio and Cavalier 1994).

SPARES AND SUPPLIES INVENTORY LEVELS AND VARIABILITY

Even if the spares and supplies dispatch stations are sufficient in number and positioned optimally, wait times will be incurred if there is insufficient inventory. The two most effective ways to determine spares and supplies inventory levels are by means of either simulation (where, again, the impact of inventory levels on factory performance must be modeled appropriately) or through the employment of analytical inventory-level models (Ignizio and Gupta, 1975; Hillier and Lieberman, 2005; Jensen and Bard, 2003; Taha, 2006). It should be noted, however, that the development of analytical models for factories with a large number of machines (and thus a large number of part and supply types) is hardly a trivial matter.

Whether you employ simulation or an analytical model, there is one mistake that you must take care to avoid. This is the matter of ignoring the importance of, and impact on wait time for, each specific type of part or supply. This is analogous to the perils of ignoring the importance of, and impact on the wait time for, the arrival of the MTs assigned to each specific workstation.

For example, if the number of spare parts stocked is based solely or mainly on historical mean-time-to-fail (MTTF) data, one might conclude that a particular part, say, part *X* of machine type A, fails so seldom that only one spare of this type needs to be stocked at any given time. Following this (questionable and indefensible) logic, one might further decide that part *Y* of machine type B fails so frequently that a large number of spares of part *Y* should be kept in inventory.

Unfortunately, this line of reasoning ignores the importance and impact of a failure of both parts *X* and *Y* on the performance of the factory. Continuing this illustration, assume that machine A is the sole machine contained within workstation A. Thus, if part *X* of machine A fails, the entire workstation will go down for repair. Further, if machine A is repaired, using the single unit of spare part *X*, and that part happens to fail before the inventory of part *X* is restocked, the impact on factory performance may be truly substantial.

Just as in the case of assigning MTs to workstations, determination of the number of spares (and amount of supplies) in support of each workstation must explicitly consider the importance of and impact on the overall performance of the factory. In fact, the general methodology discussed for the allocation of MTs to workstations—either optimally or heuristically—may be modified to encompass this situation. For the simple case (i.e., workstations consisting of identical machines), either the optimization model or the heuristic method described for MT allocation may be revised to deal with spares and supplies inventory levels. One complication to either of these approaches is the fact that storage space for spares and supplies is limited (as is the budget for purchase and storage of spares and supplies). However, this can be handled by means of constrained optimization (Ignizio and Cavalier, 1994).

An alternate approach is to employ a simulation model that incorporates spares and supplies inventory levels. After each simulation of the factory, the level of spares and supplies is revised (typically by means of a version of the greedy heuristic)—considering storage limitations—and the simulation is rerun. Continue this procedure until no further increase in factory performance seems feasible. While this approach is highly unlikely to reach even a near-optimal solution, it generally produces a "good enough" result.

Unfortunately, some firms ignore the importance and impact on factory performance of each type of spare part or supply. One such example is that of a firm that sought to reduce factory costs by reducing the number of spares kept in inventory. The spares inventory policy of the firm was to determine the average number of each type of spare part that had been required in each quarter and then multiply that number by an arbitrary weighting factor. The number of spares of each type then was established—with ad hoc consideration given to the limitations of the storage space for the inventory of spares.

After a "suggestion" by the CEO that production costs must be trimmed (including the costs of storing the inventory of spare parts), it was decided to reduce the number of *all* spare parts by one-third. More specifically, a reduction of one-third of each and every type of spare parts was dictated. It was clearly not recognized that an across-the-board reduction of spares was nonsensical. What should have been done was to prioritize parts by their importance and impact on factory performance.

Furthermore, as in many instances of factory decisions, no consideration was given to the value of cycle time. Therefore, while the cost of spares and supplies was reduced, the more important negative impact on factory performance and the firm's bottom line was substantial.

CHASING WIP BUBBLES AND VARIABILITY

Jay Forrester, in his pioneering work in systems and industrial dynamics (Forrester, 1999), demonstrated the futility of attempting to contain the oscillations of stochastic systems by means of conventional wisdom (and hasty reaction). A factory is just one type of stochastic system, and the fluctuation of its inventory levels represents one example of oscillation. Forrester, by means of simulations and feedback theory, showed that many, if not most, of the decisions taken in an attempt to stabilize a system (e.g., a factory or a supply chain) may only serve to make matters worse. This observation holds true in instances in which factory personnel engage in the ill-advised but all too common practice of chasing WIP bubbles.

A *WIP bubble* is an unanticipated increase in the number of jobs (i.e., a bubble) flowing toward a specific section of the production line. Consider, for example, a factory in which the expected number of jobs arriving at a given process step (say, process step *X*) previously has been estimated to be on the order of 10 jobs per hour. For some time that prediction has held; that is, the average arrival rate at process step *X* has been approximately 10 jobs per hour with minimal standard deviation. Unfortunately, one day several machines (which happen to have very fast process rates) in the front portion of the production line incur unexpected downtimes. Consequently (and amplified by a lack of a factory starts load-adjustment policy), a large queue forms in front of the afflicted machines. Ultimately, the machines are repaired, and as a consequence of their fast processing rate, they soon will send a much higher than usual number of jobs to the machines supporting process step *X*.

The reaction on the part of some factory managers and engineers may be to shift resources to the machines supporting process step *X*, the machines that soon will be "slammed" by the WIP bubble. For example, personnel assigned to other machines may be reassigned to support the process step *X* machines. Or if these machines support multiple process steps (e.g., a reentrant factory, such as a semiconductor wafer fabrication facility), a decision may be made to reassign the priority and dedications of the machines to process step *X*. Either way, there is a shift of resources from the support of some process steps to the step that is to soon receive the WIP bubble. Such an action, one manager informed me, is only natural. "We can't," he went on to say, "just sit back and do nothing."

Such actions, however, will only either make matters worse or simply move the impact of the WIP bubble to the machines supporting process steps farther downstream. Most likely this will increase the variability of arrival rates of machines supporting process steps downstream. And by now we know that such an increase in variability will degrade overall factory performance.

When I bring this subject up or demonstrate it by means of simulation in real-world factories, there is a predictable reaction. What, I am asked, are we supposed to do in the face of WIP bubbles? Nothing?

Frankly, doing nothing is sometimes the best course of action. If the factory has been designed properly and employs the manufacturing protocols outlined in this book, the impact on factory performance will be transient, and performance soon should return to an acceptable level. In fact, if any action is to be taken, it should be to reduce factory starts to maintain the level of inventory in the factory at a prespecified level—the level required to retain factory cycle time at the desired value (e.g., according to the first fundamental equation of manufacturing, i.e., Little's equation).

Reducing the impact of WIP bubbles and attaining a desired level of factory performance may be achieved by

- Avoiding the temptation to chase WIP bubbles by reassigning resources
- Employing appropriate manufacturing protocols, including a factory starts protocol linked to factory health and loading
- Using scientifically based WIP management (e.g., job-dispatch) rules

The latter concept is discussed in the following section.

WIP MANAGEMENT SCHEMES AND VARIABLITY

Some firms, particularly those having highly reentrant factories, such as in semiconductor manufacturing, expend enormous amounts of time, energy, and funds on the development of WIP management schemes, that is, the rules employed to dispatch jobs to the machines within a workstation. A factory starts protocol, as discussed previously, represents a type of macro-level WIP management scheme. When dealing with job-to-machine dispatch rules, the problem exists at a micro level (i.e., workstation or process-step level), and that is the matter of interest in this section.

The schemes commonly employed in WIP management at the workstation level are identical to those discussed in the literature on the general topic of sequencing and scheduling (Baker, 1974). These include first in, first out (FIFO); last in, first out (LIFO, or back-to-front); cyclic (a.k.a. *round robin*); critical ratio (CR); and on and on. Sometimes the WIP management scheme is static; that is, no matter what, the same scheme is used for the workstation. Other times the schemes are dynamic (and even frantic), changing from one set of rules to another depending on the real or perceived condition of the workstation or factory. Schemes of a dynamic nature are seen often in factories involved in chasing WIP bubbles and have the same problems as discussed in the preceding section.

WIP management (a.k.a. *job-machine dispatching*) must be employed with care. First, the type of production line must be factored into any decision made concerning a WIP management scheme. Consider, for example, a synchronous production line (e.g., a bottling plant or automobile assembly line). Such a production line is, at least for the assembly of discrete items, the closest we come to the ideal factory. There is (normally) no batching, cascading, or priority jobs. As such, the only WIP management scheme that should be required is FIFO; that is, the first job to arrive at a workstation is the first job to be selected for processing.

Another common type of production line is that which supports several types of jobs, and each job type may follow a different process-step flow through the facility. Some workstations may support several job types, whereas others may be limited to a single type. In this type of factory, customer due dates are of definite importance. Consequently, some form of the CR WIP management scheme may be appropriate.

Consider, for example, a workstation that is shared by two different job types, type A and type B. For the sake of discussion, assume that one type A job and one type B job are in the queue awaiting processing at this workstation. The decision to be made, as soon as a machine in the workstation becomes available, is which of the two jobs should be sent to it. This decision may be made by selecting the job with the smallest critical ratio.

The CR of a job is found, in turn, by dividing the time remaining (the due date of the job minus the current time) by the predicted processing time remaining for the job. Given our example with two jobs, assume that the current time is time zero. Further assume that job *A* is due 20 hours from now, whereas job *B* is due in 10 hours.

The predicted processing time remaining for job *A* is 14 hours, whereas that of job *B* is 9 hours. Table 11.2 summarizes these data and lists the associated CRs. As may be seen, job *B* has the smallest CR and should be the next job to be assigned to the next idle machine in this workstation.

While the CR WIP management scheme is often the most effective scheme for a multiproduct, multiple-process-step-flow, shared-workstation facility, high-volume production lines with few products and reentrancy are better served by another approach. This WIP management scheme, designated herein as the *minimal variability dispatch method* (MVDM), dispatches jobs so that the variability of the departures from the workstation (and subsequent variability of arrivals at the workstation or machines supporting the next process step) is minimized.

The most basic and simplest form of the MVDM WIP management scheme may be illustrated by means of an example. Assume that several jobs are in the queue in front of a workstation that supports multiple operations. Some of these jobs require one operation, whereas others require some other operation. The decision that must be made is: Which of these jobs should be sent to the next available machine in this workstation?

The job that minimizes the arrival-rate variability at the machines supporting the next process step should be the job selected for processing. One way to roughly approximate this is to select the job to be processed at this workstation that was least recently processed at its next process step. Thus, if the situation is as depicted in Table 11.3, the job to process next is job 3.

Schemes exist (somewhat more complex schemes) that reduce departure/arrival-rate variability more effectively, but this example

TABLE 11.2

Computation of CR Values

	Job *A*	**Job *B***
Due date	20 hours from now	10 hours from now
Processing time remaining	14 hours	9 hours
CR	(20 − 0)/14 = 1.43	(10 − 0)/9 = 1.11

TABLE 11.3

Minimal Variability WIP Management Scheme Approximation

Jobs in Queue	Job Type	Hours Since the Last Job of This Type Was Processed at the Next Process Step
1	*X*	3 hours
2	*X*	3 hours
3	*Y*	6 hours
4	*Z*	2 hours

should indicate the general concept. There also may be an advantage in combining the MVDM scheme with the CR approach. For example, should there be a tie in the selection of jobs for dispatch, based on minimization of variability, the tie may be broken by selecting the job having the minimum CR value.

Before leaving this topic, another WIP management scheme should be discussed. Some firms assign priorities to the jobs to be processed. Priority jobs, however, always degrade overall factory performance (Clason, 2003). A higher-priority job may jump the queue and be processed before a lower-priority job that actually has been waiting in the queue longer. It may even be the case that a workstation is kept idle, even though there are other jobs in queue, so that it is immediately available when a high-priority job is expected to arrive.

If a factory has been designed properly, and if the appropriate manufacturing protocols have been implemented, there should be no need to employ priority jobs. They only serve to complicate matters, induce variability, and degrade factory cycle time for the nonpriority jobs. The larger the volume of priority jobs, the worse is the situation.

Consider, for example, a facility that originally had priority-job cycle times of 20 days and non-priority-job cycle times of 90 days. By implementing enhanced manufacturing protocols and eliminating the prioritization of jobs, the cycle time for all jobs was reduced to 22 days. As such, the average cycle time of any job was almost the same as that of the priority jobs in the factory's original configuration.

To repeat, avoid whenever possible the prioritization of jobs in high-volume reentrant factories, and in general, never allow the volume of priority jobs in such a facility to exceed 5 percent of the total.

CHAPTER SUMMARY

Complexity and variability are two of the three enemies of factory performance. Chapter 10 presented guidelines for the reduction of complexity. This chapter introduced guidelines useful for reducing variability. Chapter 12 provides a brief overview of the implementation of these guidelines by means of a return to the 12-workstation model.

CASE STUDY 11: HURRY UP . . . AND PAY THE PRICE

Tommy issued several follow-on orders in an attempt to bolster his “Hurry-Up” campaign. All leaves, with the exception of those for pregnancy, and all vacations were canceled for factory floor personnel. To further increase factory output, he demanded that all PM events and efforts be halted. “Run,” Tommy said, “until the machines break down, and then fix them.”

The “Hurry-Up” effort appeared to work, at least at first. Frantic factory floor personnel managed to up the factory outs to almost 11,000 units a week. Marvin Muddle seems placated, at least for the moment. When it seemed that the capacity problem had been licked, Tommy held a rally in the complex’s cafeteria. Everyone was given his or her choice of a free ice cream sandwich or soft drink.

Unfortunately for Tommy, now—two months later—chaos reigns. And Tommy seems to be receiving the full force of the blame.

* * *

“Have you seen the latest figures on factory performance,” says Dan. “It seems that our cycle time is worse than it has ever been, and factory inventory is enormous. They can’t find anyplace to store the in-process work. That ‘Hurry-Up’ program has destroyed the factory and antagonized most of the floor personnel. From what I hear, the best people have either left or are looking for other jobs. What a mess!”

“If you recall,” says Julia, “that’s precisely what Professor Leonidas said would happen. His exact words were, if I recall correctly, ‘You can only ask so much of people and machines. Anyone foolish enough to try to defy the laws of physics and human nature is begging for trouble.' And trouble is exactly what we are seeing."

"I talked to one of my friends, a factory floor supervisor for a workstation in the front end of the production line," says Brad. "He told me that the machines are in terrible shape. As a result of putting off their PM events, a number of critical and costly machines were ruined. Others are off-line while the maintenance and repair crews do what they can to get them back up and running. But I think that the 'Hurry-Up' program broke one thing that will never be fixed—the trust and morale of the people who had to endure this short-sighted campaign."

"I even heard that some of the members of the *LEAN* Forward team have left," says Dan. "And there was even a fistfight between a couple of their people and some members of the quality control team. But hey, what the heck, I did get a free ice cream sandwich."

"People," says Brad, "I've got an announcement to make. I hate to leave you all in the lurch, but I've had enough. I see no point in trying to get Muddle interested in science. This place is run by the seat-of-the-pants, and that's never going to change. I turned in my resignation this morning. Two weeks from now I'll no longer be an employee of Muddle. I wish you all the best, but I just can't take it anymore."

"I'm sorry to hear that," says Julia. "Can you tell us where you are going? Do they need anyone with my background? And I mean that seriously."

"Sally Swindel and I have decided to pursue, as they say, other interests. We intend to form our own management consulting firm once we return from our honeymoon."

"Congratulations, Brad," says Dan, trying to contain his shock. "What type of management consulting will you two be doing, if I may ask?"

"You may. Sally and I came up with an idea for a book on leadership some time ago. I've put together a rough draft and hope to have the book finished in a few months. Once that's done, Sally is going to do her thing. She's going to market the book and the associated training courses. As you know, she can be pretty persuasive."

"What's the title of the book?" asks Winston.

"The tentative title is *Leadership Principles of the Donner Party: How to Overcome Any Obstacle.*"

"The Donner party?" asks Dan. "Isn't that the group of emigrants who sought their fortune in the West and tried to make their way to California? Isn't that the same group of pinheads that got caught in a blizzard and resorted to cannibalism? Didn't about half those poor souls die?"

"You've got it," says Brad, grinning broadly. "Sally and I are convinced that the book will be a best-seller and that we can charge top prices for our training courses."

Winston could only wonder if those training courses would take place in the dead of winter without anything to eat but your fellow attendees. I guess, Winston thinks, that I will never understand the management consulting business.

CHAPTER 11 EXERCISES

1. Batching induces variability in a factory. Suggestions to reduce batch sizes, however, are met with resistance. The arguments made for not reducing batch sizes are
    - If batch sizes are reduced, the full capacity of the batching machines will not be exploited.
    - Smaller batch sizes produce the need for more setups.

   These arguments against smaller batch sizes are valid. So why might smaller batch sizes still improve *overall* factory performance, and how could you prove it?
2. The arguments against declustering of factory starts are that declustering will require more oversight as well as more trips to the factory starts site. What is your counterargument?
3. How do clustered and minimal overlap PM activities induce variability?
4. You have developed an optimized MT allocation scheme for a factory, but it is met with resistance on the part of the factory floor supervisor for the workstation that has, in the past, always been the factory constraint. The machines in that workstation are also by far the most expensive in the factory. In the past, 25 percent of the MTs on each shift were allocated to this workstation. Your optimal plan reduces that proportion to just 20 percent. The floor supervisor demands a reallocation based on the traditional allocation rate. What is your counterargument?
5. Whether it is for spare parts or supplies, what is your counterargument for those who believe that if any reductions are to be made, they must be across the board (e.g., "equal pain")?

6. In one firm there are daily operations meetings in which the state of the factory is discussed. The primary output of these meetings is recommendations for the shift of resources from one workstation or set of machines to another based on predictions of incoming WIP bubbles. Explain why such frequent changes are usually counterproductive. Explain why such a practice is so attractive to factory management and engineers.

CHAPTER 12

The 12-Workstation Model Revisited

Several of the guidelines for improving factory performance (i.e., described in the preceding two chapters) may be illustrated by means of a revised version of the 12-workstation factory. In this form of the model, our primary objective is to maximize profit rather than simply minimize factory cycle time.

THE ATTRIBUTES OF THE FACTORY

The basic configuration of the revised 12-workstation factory is similar to that of the models presented in Chapters 4 and 6. The only significant differences are as follows:

- Workstation D now has, in its initial state, 6 machines rather than 5.
- Workstation F now has, in its initial state, 12 machines rather than 10.
- The effective process rates (EPRs) of the machines in each workstation have been separated into two components: their raw process rates and the availability achieved via the choice of the number of maintenance technicians allocated to the workstations.
- Every day that factory cycle time decreases below that attained in the initial scenario increases profit by $1M.

The objective of this exercise is to allocate funding to (1) purchase additional machines, (2) increase the raw process rates of the machines, (3) increase or decrease the number of maintenance technicians (MTs) allocated to each workstation, or (4) reduce the variability imposed by either the factory starts protocol or effective process times of the workstations—and do so to maximize profit (i.e., rather than just minimize factory cycle time, as was the objective in Chapters 4 and 6).

Figure 12.1 presents the revised workstation-centric flowchart for the 12-workstation factory. As before, there are zero transit times between workstations and no reentrancy or scrap. The blocks within each workstation, as before, indicate the number of machines that exist initially in the associated workstation. The arrows show direction of job flow from workstation to workstation.

Assuming, as before, that every machine in a given workstation is qualified to support the (single) process step conducted by that workstation, an equivalent process-step-centric model may be constructed for the 12-workstation factory. That model is shown in Figure 12.2. In that figure, the machines supporting each process step are listed in the triangle under the associated process step. For example, process step 2 is supported by machines B1, B2, and B3 (i.e., B1 through B3, designated in the figure as B1–B3) of workstation B.

The numbers in parentheses above each transit-step arrow indicate that the throughput flow rate of jobs through the factory (and through each workstation in the factory) is—as in the original model—an average of 20 units per day. Additional details about the attributes of the factory are presented in the next section.

FIGURE 12.1

Workstation-centric flowchart for revised 12-workstation factory.

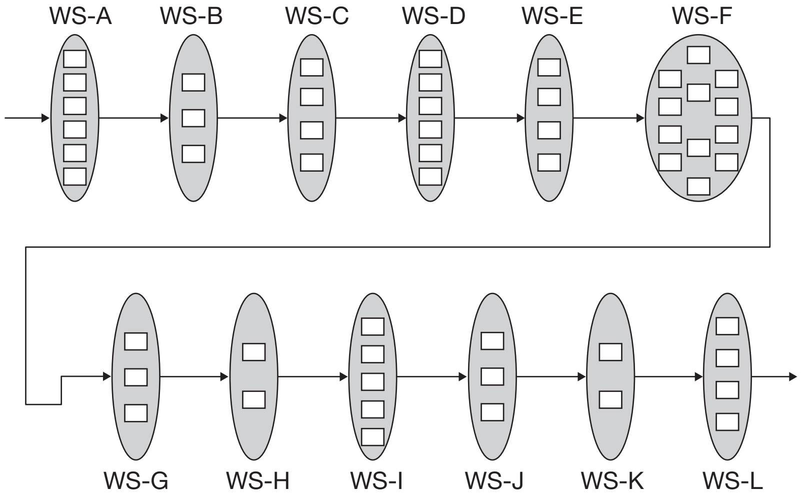
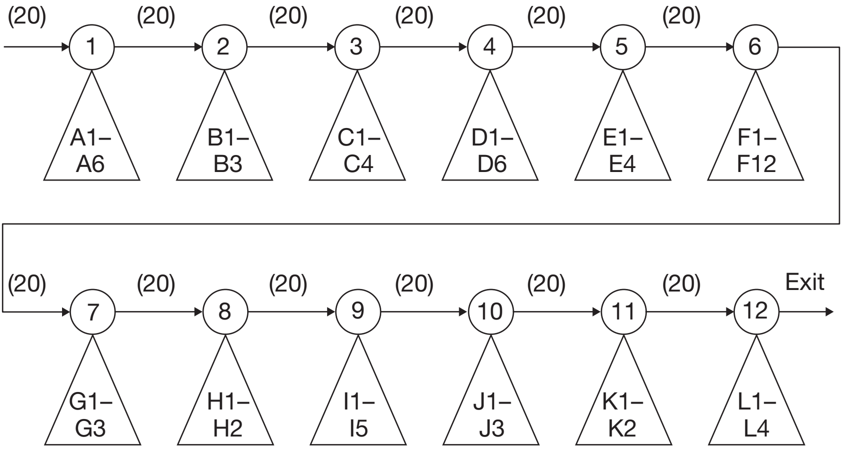


FIGURE 12.2

Process-step-centric flowchart for revised 12-workstation factory.



PROBLEM STATEMENT

Presently, the cycle time of the factory is 211.76 days—which is assumed to be much, much worse than that of your competition (and even worse than the initial factory configurations of Chapters 4 and 6). Now, however, maximizing profit is your primary goal (and reduction of factory cycle time is secondary).

Your mission is to maximize profit by means of one or more of the following actions:

- Allocate the funds necessary to add additional machines to one or more of the workstations.
- Increase or reduce the MTs assigned to each workstation (which, in turn, affects availability), where the cost of an MT is assumed to be \$1M (\$100,000) each.
- Allocate funds to increase the raw process rates (*PR*) of the machines in a workstation.
- Allocate funds to decrease the primary sources of factory variability (i.e., the coefficient of variability values determined by either factory starts or effective process times).
- Or use some combination of the preceding.

Any of these alternatives involves money, and for sake of discussion, we will assume that our total budget is limited to \$15M.

Figure 12.3 serves to summarize, in virtually the same matrix form as employed earlier, the present condition of the factory. Before proceeding further, however, we should discuss the impact of the number of MTs on the availability (and subsequent *EPR*) of the workstation.

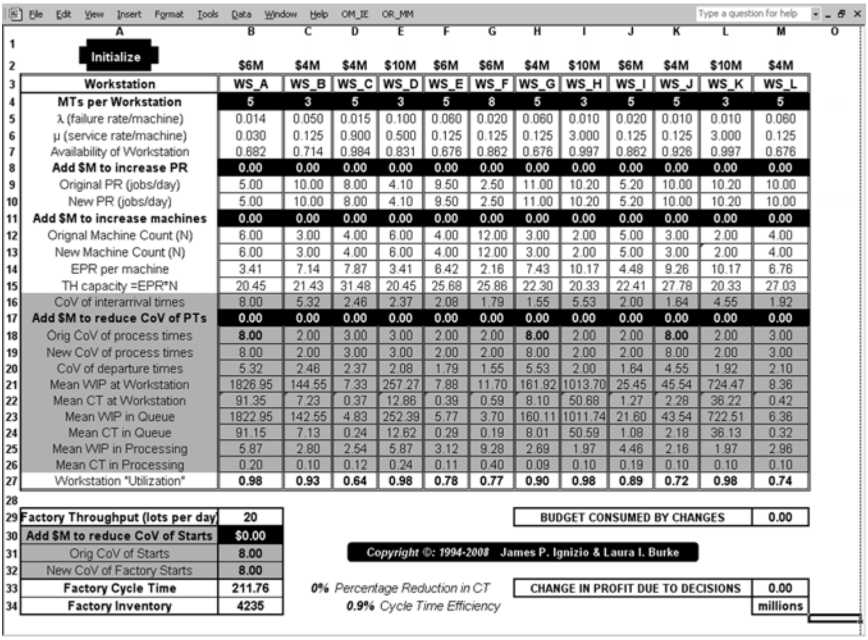
Row 7, cells B7 through M7 list the availability of each workstation as a function of the number of MTs assigned in cells B4 through M4 and subject to failure rate λ and repair rate μ values in rows 5 and 6 (cells B5:M5 and B6:M6). For example, the failure rate of the machines in workstation A is 0.014 (i.e., 0.014 failures per machine per hour). The inverse of this value is 1/0.014, or 71.4 hours, and represents the mean time to failure (MTTF).

The repair (or service) rate of the machines in workstation A is shown as 0.030 (i.e., the average service rate is 0.030 machines per hour). The inverse of this value is 1/0.030, or 33.33 hours, and is the mean time to repair (MTTR) for each machine in workstation A.

Given the number of MTs (e.g., servers) and the arrival-rate and service-rate values, the queuing theory may be used to compute the

FIGURE 12.3

Twelve-workstation model for profit maximization, initial scenario.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	Initialize	$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
3	**Workstation**	**WS_A**	**WS_B**	**WS_C**	**WS_D**	**WS_E**	**WS_F**	**WS_G**	**WS_H**	**WS_I**	**WS_J**	**WS_K**	**WS_L**
4	**MTs per Workstation**	**5**	**3**	**5**	**3**	**5**	**8**	**5**	**3**	**5**	**5**	**3**	**5**
5	λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
6	µ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
7	Availability of Workstation	0.682	0.714	0.984	0.831	0.676	0.862	0.676	0.997	0.862	0.926	0.997	0.676
8	**Add $M to increase PR**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
9	Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
10	New PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
11	**Add $M to increase machines**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
13	New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
14	EPR per machine	3.41	7.14	7.87	3.41	6.42	2.16	7.43	10.17	4.48	9.26	10.17	6.76
15	TH capacity =EPR*N	20.45	21.43	31.48	20.45	25.68	25.86	22.30	20.33	22.41	27.78	20.33	27.03
16	CoV of interarrival times	8.00	5.32	2.46	2.37	2.08	1.79	1.55	5.53	2.00	1.64	4.55	1.92
17	**Add $M to reduce CoV of PTs**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
18	Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
19	New CoV of process times	8.00	2.00	3.00	3.00	2.00	2.00	8.00	2.00	2.00	8.00	2.00	3.00
20	CoV of departure times	5.32	2.46	2.37	2.08	1.79	1.55	5.53	2.00	1.64	4.55	1.92	2.10
21	Mean WIP at Workstation	1826.95	144.55	7.33	257.27	7.88	11.70	161.92	1013.70	25.45	45.54	724.47	8.36
22	Mean CT at Workstation	91.35	7.23	0.37	12.86	0.39	0.59	8.10	50.68	1.27	2.28	36.22	0.42
23	Mean WIP in Queue	1822.95	142.55	4.83	252.39	5.77	3.70	160.11	1011.74	21.60	43.54	722.51	6.36
24	Mean CT in Queue	91.15	7.13	0.24	12.62	0.29	0.19	8.01	50.59	1.08	2.18	36.13	0.32
25	Mean WIP in Processing	5.87	2.80	2.54	5.87	3.12	9.28	2.69	1.97	4.46	2.16	1.97	2.96
26	Mean CT in Processing	0.20	0.10	0.12	0.24	0.11	0.40	0.09	0.10	0.19	0.10	0.10	0.10
27	Workstation "Utilization"	**0.98**	**0.93**	**0.64**	**0.98**	**0.78**	**0.77**	**0.90**	**0.98**	**0.89**	**0.72**	**0.98**	**0.74**
28													

29	**Factory Throughput (lots per day)**	**20**
30	**Add $M to reduce CoV of Starts**	**$0.00**
31	Orig CoV of Starts	**8.00**
32	New CoV of Factory Starts	**8.00**
33	**Factory Cycle Time**	**211.76**
34	**Factory Inventory**	**4235**

BUDGET CONSUMED BY CHANGES	**0.00**
CHANGE IN PROFIT DUE TO DECISIONS	**0.00**
	millions

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

0% *Percentage Reduction in CT*
0.9% *Cycle Time Efficiency*

workstation availability values (Hillier and Lieberman, 2005; Taha, 2006). This is precisely what has been used in the figure. The availability value thus computed then is used to determine the *EPR* of each machine and the throughput capacity (*TH*) of each workstation. Other than this, the 12-workstation model performs in an identical fashion to that employed in Chapter 6.

PROBLEM SOLUTION

The objective of the solution process is to perform resource allocation and reallocation in a manner that maximizes profit. Readers are encouraged to attempt to accomplish this task before proceeding further. The spreadsheet for the model may be found at

www.mhprofessional.com/Ignizio/12WS_Ch12

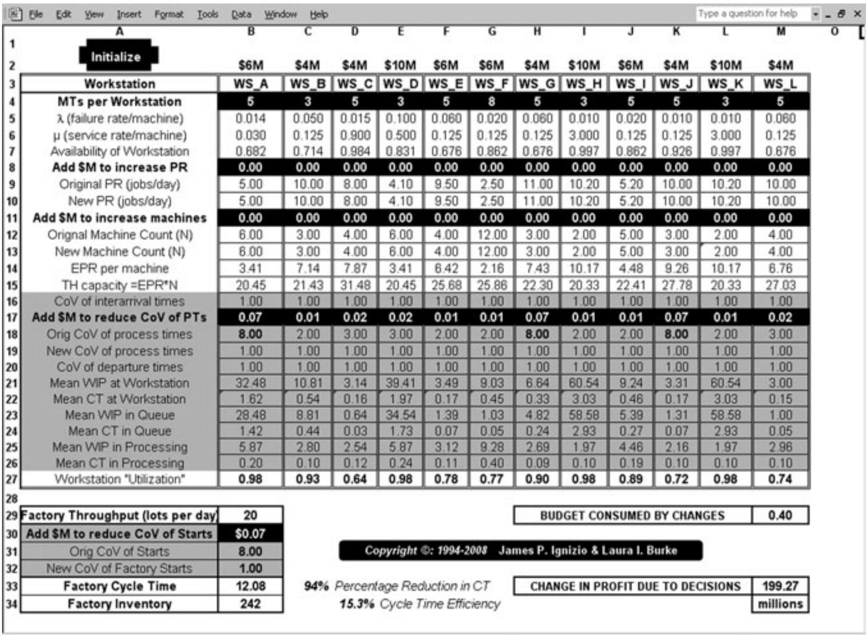
Once you have completed this exercise, proceed to the next paragraph and compare your results with those achieved here.

If we apply the guidelines indicated in Chapters 10 and 11, our first step (phase 1) is to allocate funds to reduce variability. As in Chapter 6, we assume that we will attempt to reduce the coefficient of variability (*CoV*) values of factory starts and process times to values of 1.0 or until funds run out. Priority is given to reducing the *CoV* values that are the largest and closest to the factory input. The result of funding to reduce variability is presented in Figure 12.4. Note that with an expenditure of $0.4M ($400,000), we are able to reduce all *CoV* values to 1.0 with an associated value for profit and factory cycle time of $199.27M and 12.08 days, respectively. This represents a 94 percent reduction in cycle time. All this was accomplished simply (and cheaply) by reducing factory variability!

Our next step (phase 2) is to reallocate the MTs assigned to each workstation. While this could be accomplished by means of optimization (Ignizio, 2004; Ignizio and Cavalier, 1994), we will follow the heuristic guidelines from Chapter 11. More specifically, priorities will be established via the heuristic introduced in Chapter 11, and using these, the MTs will be reallocated. Note carefully, however, that if the number of MTs assigned to a workstation is insufficient to avoid development of an infinite queue, the cell associated with workstation utilization will turn red. In such an instance, increase the number of MTs for that workstation until the color of that cell is again white.

FIGURE 12.4

Twelve-workstation model for profit maximization, phase 1.



	$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
Workstation	WS_A	WS_B	WS_C	WS_D	WS_E	WS_F	WS_G	WS_H	WS_I	WS_J	WS_K	WS_L
MTs per Workstation	5	3	5	3	5	8	5	3	5	5	3	5
λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
μ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
Availability of Workstation	0.682	0.714	0.984	0.831	0.676	0.862	0.676	0.997	0.862	0.926	0.997	0.676
Add $M to increase PR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
New PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
Add $M to increase machines	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
EPR per machine	3.41	7.14	7.87	3.41	6.42	2.16	7.43	10.17	4.48	9.26	10.17	6.76
TH capacity =EPR*N	20.45	21.43	31.48	20.45	25.68	25.86	22.30	20.33	22.41	27.78	20.33	27.03
CoV of interarrival times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Add $M to reduce CoV of PTs	0.07	0.01	0.02	0.02	0.01	0.01	0.07	0.01	0.01	0.07	0.01	0.02
Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
New CoV of process times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CoV of departure times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Mean WIP at Workstation	32.48	10.81	3.14	39.41	3.49	9.03	6.64	60.54	9.24	3.31	60.54	3.00
Mean CT at Workstation	1.62	0.54	0.16	1.97	0.17	0.45	0.33	3.03	0.46	0.17	3.03	0.15
Mean WIP in Queue	28.48	8.81	0.64	34.54	1.39	1.03	4.82	58.58	5.39	1.31	58.58	1.00
Mean CT in Queue	1.42	0.44	0.03	1.73	0.07	0.05	0.24	2.93	0.27	0.07	2.93	0.05
Mean WIP in Processing	5.87	2.80	2.54	5.87	3.12	9.28	2.69	1.97	4.46	2.16	1.97	2.96
Mean CT in Processing	0.20	0.10	0.12	0.24	0.11	0.40	0.09	0.10	0.19	0.10	0.10	0.10
Workstation "Utilization"	**0.98**	**0.93**	**0.64**	**0.98**	**0.78**	**0.77**	**0.90**	**0.98**	**0.89**	**0.72**	**0.98**	**0.74**

Factory Throughput (lots per day)	20
Add $M to reduce CoV of Starts	$0.07
Orig CoV of Starts	8.00
New CoV of Factory Starts	1.00
Factory Cycle Time	12.08
Factory Inventory	242

BUDGET CONSUMED BY CHANGES	0.40
CHANGE IN PROFIT DUE TO DECISIONS	199.27
	millions

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

94% *Percentage Reduction in CT*
15.3% *Cycle Time Efficiency*

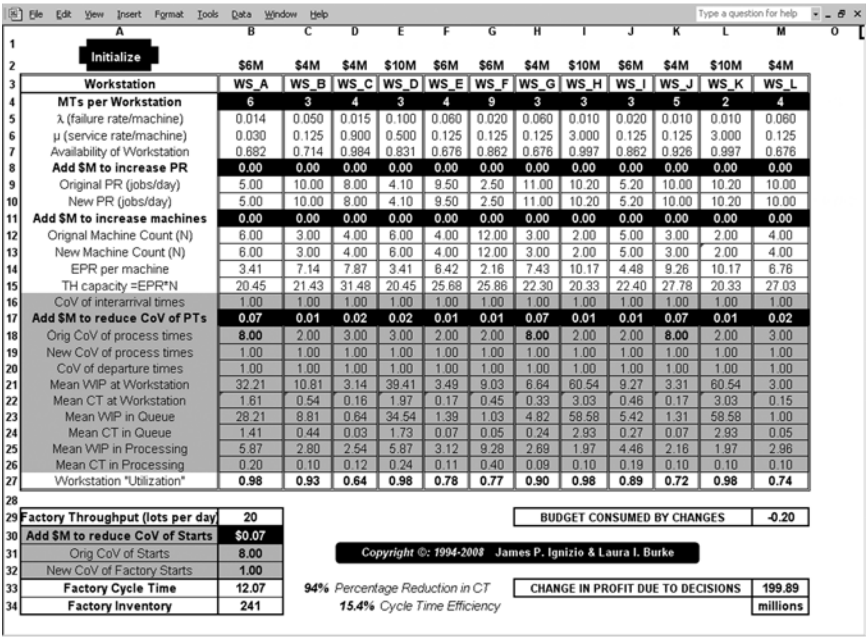
One of several alternative reallocations is shown in Figure 12.5. Notice that the total number of MTs has been reduced from 55 to 49. The subsequent profit and factory cycle time has been changed slightly to $199.89M and 12.07 days, respectively. We have, however, reduced the cost of the changes to −$0.2M (i.e., a savings of $200,000 over the original factory configuration).

The third phase of the effort is to allocate funds for either adding machines or increasing raw process rates. If we prioritize the workstations that are either constraints or that feed constraints, one possible allocation of funds produces the matrix shown in Figure 12.6. In this final phase of the heuristic process, the profit has increased to $205.72M, whereas factory cycle time has been reduced to just 3.74 days (and a quite impressive *CTE* of 47 percent).

Our result, obtained by means of the heuristic guidelines (depending on tie-breaking rules, even better solutions are possible), may be compared with that found by optimization (i.e., via the first phase of a genetic algorithm). This optimal solution is shown in Figure 12.7. It may be noted that the optimal solution has a profit of only a little more than 1 percent greater than that

FIGURE 12.5

Twelve-workstation model for profit maximization, phase 2.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Initialize												
2		$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
3	Workstation	WS_A	WS_B	WS_C	WS_D	WS_E	WS_F	WS_G	WS_H	WS_I	WS_J	WS_K	WS_L
4	**MTs per Workstation**	6	3	4	3	4	9	3	3	3	5	2	4
5	λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
6	μ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
7	Availability of Workstation	0.682	0.714	0.984	0.831	0.676	0.862	0.676	0.997	0.862	0.926	0.997	0.676
8	**Add $M to increase PR**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
10	New PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
11	**Add $M to increase machines**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
13	New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
14	EPR per machine	3.41	7.14	7.87	3.41	6.42	2.16	7.43	10.17	4.48	9.26	10.17	6.76
15	TH capacity =EPR*N	20.45	21.43	31.48	20.45	25.68	25.86	22.30	20.33	22.40	27.78	20.33	27.03
16	CoV of interarrival times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	**Add $M to reduce CoV of PTs**	0.07	0.01	0.02	0.02	0.01	0.01	0.07	0.01	0.01	0.07	0.01	0.02
18	Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
19	New CoV of process times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	CoV of departure times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21	Mean WIP at Workstation	32.21	10.81	3.14	39.41	3.49	9.03	6.64	60.54	9.27	3.31	60.54	3.00
22	Mean CT at Workstation	1.61	0.54	0.16	1.97	0.17	0.45	0.33	3.03	0.46	0.17	3.03	0.15
23	Mean WIP in Queue	28.21	8.81	0.64	34.54	1.39	1.03	4.82	58.58	5.42	1.31	58.58	1.00
24	Mean CT in Queue	1.41	0.44	0.03	1.73	0.07	0.05	0.24	2.93	0.27	0.07	2.93	0.05
25	Mean WIP in Processing	5.87	2.80	2.54	5.87	3.12	9.28	2.69	1.97	4.46	2.16	1.97	2.96
26	Mean CT in Processing	0.20	0.10	0.12	0.24	0.11	0.40	0.09	0.10	0.19	0.10	0.10	0.10
27	Workstation "Utilization"	**0.98**	**0.93**	**0.64**	**0.98**	**0.78**	**0.77**	**0.90**	**0.98**	**0.89**	**0.72**	**0.98**	**0.74**
28													
29	**Factory Throughput (lots per day)**	20											
30	**Add $M to reduce CoV of Starts**	$0.07											
31	Orig CoV of Starts	**8.00**											
32	New CoV of Factory Starts	**1.00**											
33	**Factory Cycle Time**	12.07											
34	**Factory Inventory**	241											

BUDGET CONSUMED BY CHANGES -0.20

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

94% *Percentage Reduction in CT*
15.4% *Cycle Time Efficiency*

CHANGE IN PROFIT DUE TO DECISIONS 199.89 millions

achieved by the heuristic method. Cycle time and *CTE* for the optimal solution are also only a bit better. What is of most interest, however, is that by means of following the guidelines laid out previously, a simple heuristic has obtained a solution very close to optimal.

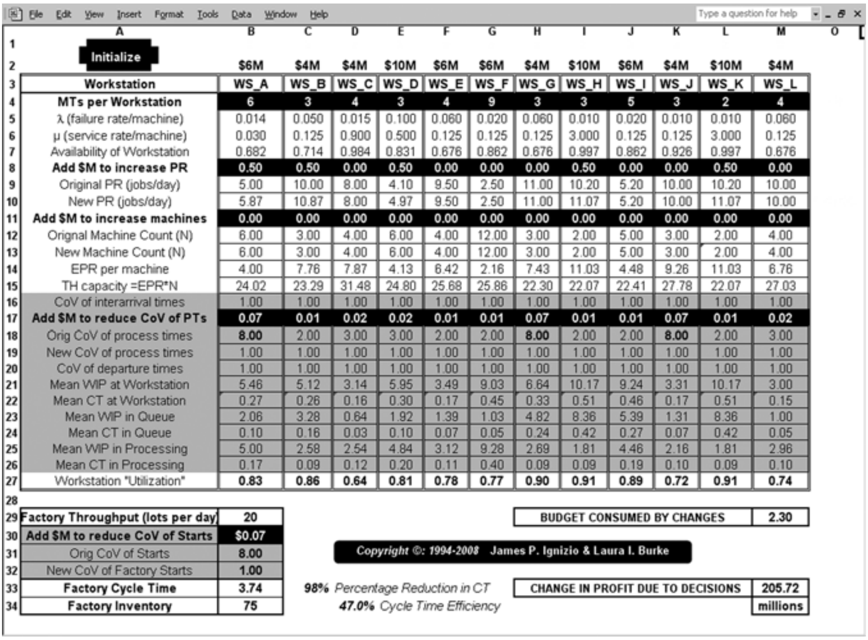
Now that we have the solution to the problem by means of either heuristic guidelines or optimization, we should examine the mechanism of the heuristic approach in more detail. For example, just how do we reduce factory variability in phase 1?

HEURISTIC PROCESS IN DETAIL

The initial configuration of the 12-workstation factory was provided in Figure 12.3. The first phase of the heuristic approach is that of reducing the inherent variability of the factory by adding funds to projects that should reduce the *CoV* of factory starts and effective process times. It is a general rule that the emphasis in either reducing variability or increasing effective capacity should be placed on a change of protocols in the front end of the

FIGURE 12.6

Twelve-workstation model for profit maximization, phase 3.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
3	Workstation	WS_A	WS_B	WS_C	WS_D	WS_E	WS_F	WS_G	WS_H	WS_I	WS_J	WS_K	WS_L
4	**MTs per Workstation**	6	3	4	3	4	9	3	3	5	3	2	4
5	λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
6	μ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
7	Availability of Workstation	0.682	0.714	0.984	0.831	0.676	0.862	0.676	0.997	0.862	0.926	0.997	0.676
8	**Add $M to increase PR**	0.50	0.50	0.00	0.50	0.00	0.00	0.00	0.50	0.00	0.00	0.50	0.00
9	Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
10	New PR (jobs/day)	5.87	10.87	8.00	4.97	9.50	2.50	11.00	11.07	5.20	10.00	11.07	10.00
11	**Add $M to increase machines**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
13	New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
14	EPR per machine	4.00	7.76	7.87	4.13	6.42	2.16	7.43	11.03	4.48	9.26	11.03	6.76
15	TH capacity =EPR*N	24.02	23.29	31.48	24.80	25.68	25.86	22.30	22.07	22.41	27.78	22.07	27.03
16	CoV of interarrival times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	**Add $M to reduce CoV of PTs**	0.07	0.01	0.02	0.02	0.01	0.01	0.07	0.01	0.01	0.07	0.01	0.02
18	Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
19	New CoV of process times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	CoV of departure times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21	Mean WIP at Workstation	5.46	5.12	3.14	5.95	3.49	9.03	6.64	10.17	9.24	3.31	10.17	3.00
22	Mean CT at Workstation	0.27	0.26	0.16	0.30	0.17	0.45	0.33	0.51	0.46	0.17	0.51	0.15
23	Mean WIP in Queue	2.06	3.28	0.64	1.92	1.39	1.03	4.82	8.36	5.39	1.31	8.36	1.00
24	Mean CT in Queue	0.10	0.16	0.03	0.10	0.07	0.05	0.24	0.42	0.27	0.07	0.42	0.05
25	Mean WIP in Processing	5.00	2.58	2.54	4.84	3.12	9.28	2.69	1.81	4.46	2.16	1.81	2.96
26	Mean CT in Processing	0.17	0.09	0.12	0.20	0.11	0.40	0.09	0.09	0.19	0.10	0.09	0.10
27	Workstation "Utilization"	**0.83**	**0.86**	**0.64**	**0.81**	**0.78**	**0.77**	**0.90**	**0.91**	**0.89**	**0.72**	**0.91**	**0.74**
28													
29	**Factory Throughput (lots per day)**	20											
30	**Add $M to reduce CoV of Starts**	$0.07											
31	Orig CoV of Starts	**8.00**											
32	New CoV of Factory Starts	**1.00**											
33	**Factory Cycle Time**	**3.74**											
34	**Factory Inventory**	**75**											

BUDGET CONSUMED BY CHANGES **2.30**

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

98% *Percentage Reduction in CT*
47.0% *Cycle Time Efficiency*

CHANGE IN PROFIT DUE TO DECISIONS **205.72**
millions

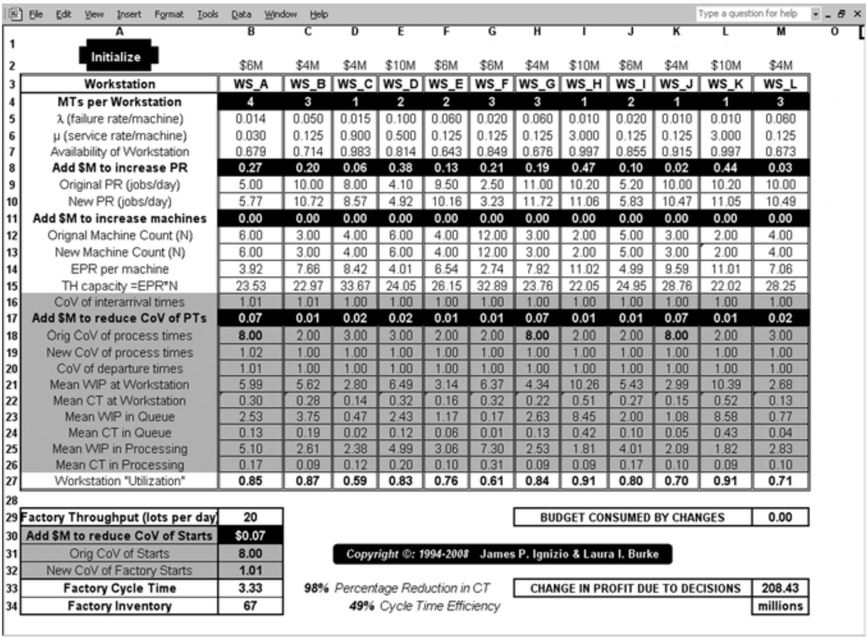
production line. As such, the first source of variability that should be dealt with is factory starts. We may devote funding up to $0.07M for reducing that CoV from its present value of 8.0 to the desired value of 1.0.

Reducing factory starts variability may be achieved by using a declustered factory starts protocol (coupled with, where possible, a factory loading scheme synchronized with factory health). Once this has been achieved, we move to the remaining largest sources of variability, with priority again given to the sources closest to the factory input. Thus we deal with the variability of the effective process time of workstation A (devoting $0.07M to that effort).

The source of variability within an effective process time of a workstation may be actual variability in the raw process rate of the machines in the workstation—a matter that may require physical changes—or, more likely, problems with operating or maintenance protocols. If the workstation requires manual operation and the presence of human operators, the cause of the variability may be too few operators, poorly trained operators, or operating specifications that are not C^4U-compliant. It is more likely (particularly in

FIGURE 12.7

Optimal solution.



	$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
Workstation	**WS_A**	**WS_B**	**WS_C**	**WS_D**	**WS_E**	**WS_F**	**WS_G**	**WS_H**	**WS_I**	**WS_J**	**WS_K**	**WS_L**
MTs per Workstation	4	3	1	2	2	3	3	1	2	1	1	3
λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
μ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
Availability of Workstation	0.679	0.714	0.983	0.814	0.643	0.849	0.676	0.997	0.855	0.915	0.997	0.673
Add $M to increase PR	0.27	0.20	0.06	0.38	0.13	0.21	0.19	0.47	0.10	0.02	0.44	0.03
Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
New PR (jobs/day)	5.77	10.72	8.57	4.92	10.16	3.23	11.72	11.06	5.83	10.47	11.05	10.49
Add $M to increase machines	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
EPR per machine	3.92	7.66	8.42	4.01	6.54	2.74	7.92	11.02	4.99	9.59	11.01	7.06
TH capacity =EPR*N	23.53	22.97	33.67	24.05	26.15	32.89	23.76	22.05	24.95	28.76	22.02	28.25
CoV of interarrival times	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Add $M to reduce CoV of PTs	0.07	0.01	0.02	0.02	0.01	0.01	0.07	0.01	0.01	0.07	0.01	0.02
Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
New CoV of process times	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CoV of departure times	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Mean WIP at Workstation	5.99	5.62	2.80	6.49	3.14	6.37	4.34	10.26	5.43	2.99	10.39	2.68
Mean CT at Workstation	0.30	0.28	0.14	0.32	0.16	0.32	0.22	0.51	0.27	0.15	0.52	0.13
Mean WIP in Queue	2.53	3.75	0.47	2.43	1.17	0.17	2.63	8.45	2.00	1.08	8.58	0.77
Mean CT in Queue	0.13	0.19	0.02	0.12	0.06	0.01	0.13	0.42	0.10	0.05	0.43	0.04
Mean WIP in Processing	5.10	2.61	2.38	4.99	3.06	7.30	2.53	1.81	4.01	2.09	1.82	2.83
Mean CT in Processing	0.17	0.09	0.12	0.20	0.10	0.31	0.09	0.09	0.17	0.10	0.09	0.10
Workstation "Utilization"	**0.85**	**0.87**	**0.59**	**0.83**	**0.76**	**0.61**	**0.84**	**0.91**	**0.80**	**0.70**	**0.91**	**0.71**

Factory Throughput (lots per day)	**20**
Add $M to reduce CoV of Starts	**$0.07**
Orig CoV of Starts	**8.00**
New CoV of Factory Starts	**1.01**
Factory Cycle Time	**3.33**
Factory Inventory	**67**

BUDGET CONSUMED BY CHANGES	**0.00**
CHANGE IN PROFIT DUE TO DECISIONS	**208.43**
	millions

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

98% *Percentage Reduction in CT*
49% *Cycle Time Efficiency*

highly automated factories), however, that the source of variability in effective process times is inferior maintenance protocols. This calls for an examination of the preventive maintenance (PM) specifications, development of an availability profile plot, and computation of the M-ratio.

Consequently, to reduce the *CoV* of the effective process times of workstation A, we may have to develop C^{4}U-compliant PM specs, decluster maintenance events, evaluate the location and levels of spares and supplies, or reallocate MTs (the latter is accomplished via the second phase of the heuristic). In short, a Waddington analysis should be considered.

Once we have reduced the *CoV* of the effective process times of workstation A, we move on to workstations G, J, C, D, L, B, E, F, H, I, and K in that order. The same approach to *CoV* reduction as described for workstation A applies to these workstations.

Before proceeding to phase 2, it should be pointed out that the 12-workstation model does not include batching and assumes that the only source of variability, other than factory starts, is effective process times. In an actual factory, we also should consider the variability caused by job arrivals at each workstation and machine and allocate the resources necessary to reduce the variability about arrivals to a reasonable level. For example, batching at a preceding process step will induce variability in the arrival rate at the next process step. Steps thus should be taken to reduce batch sizes whenever possible.

Another matter that is not dealt with in the 12-workstation factory model is that of "victim" and "villain" workstations. In a real-world factory, a specific workstation (or set of machines supporting a specific process step) might be perceived as a problem because of a higher than expected cycle time. It is all too common to jump to the conclusion that resources should be allocated to the (apparently) poorly performing workstation when, in actuality, it is the victim of a workstation supporting the preceding process step (or steps).

Remember the third fundamental equation of manufacturing, the propagation of variability. It may well be that the feeder workstation is delivering jobs to the next workstation with a high level of arrival-rate variability. This means that it is the workstation to which resources should be allocated rather than the victim workstation. While this is not dealt with in the 12-workstation factory model, it is a matter that always should be considered in practice.

Phase 2 of the heuristic approach involves the reallocation of MTs as well as a possible increase or decrease in total MT headcount. While the optimization procedure may be used here, let's restrict our discussion to implementation of the heuristic guidelines of Chapter 11. Rather, however, than computing the weighting factors and queues, as was described there, an even simpler approach may be employed. While not as effective, this simpler method is often "good enough." The approach may be summarized as follows:

Observe, at all times, the impact on profit (cell M29) of any action taken. If the action reduces profit, do not take it. Observe, at all times, the color of the cells in row 27 (cells B27 to M27). If an action taken (e.g., reducing the number of MTs at a workstation) results in a cell turning red, do not take that action.

If the number of MTs exceeds the number of machines in the workstation (and under the assumption that only one MT is required for an event), reduce that number to the number of machines. (Thus the number of MTs assigned to workstation C may be reduced to a value of 4, those to E to 4, those to G to 3, those to J to 3, those to K to 2, and those to L to 4.)

In the next step, allocate additional MTs (up to the number of machines in the workstation—again under the assumption that only one MT is required for an event) to the workstations having the highest utilization (workstations A, D, H, and K). If, however, the addition of MTs fails to reduce the utilization value (i.e., occupation rate), return the number of MTs to the original value. For this factory, at this step, additional MTs at any workstation will not improve profit.

We next attempt to reduce the number of MTs at each workstation. If the reduction improves profit while not turning a cell in row 27 red, then take that action.

Following the reallocation of MTs, we move to phase 3. In phase 3, we employ a theory of constraints–influenced approach and allocate funds to increase the capacity of constraint and near-constraint workstations, with priority given to the constraints closest to factory input. Since increasing the process rate of a machine (e.g., by modifying its physical components) is usually cheaper than adding machines, priority normally should be given to process-rate increases.

If the allocation of funds to increasing process rates or machines increases total profit, take that action. Otherwise, return to the previous factory state.

At the conclusion of these three phases, a solution close to the optimal solution (in terms of profit) should have been achieved. The result, for the steps just outlined, is even better in terms of profit (although worse in terms of cycle time and *CTE*) than was obtained previously. This is shown in Figure 12.8.

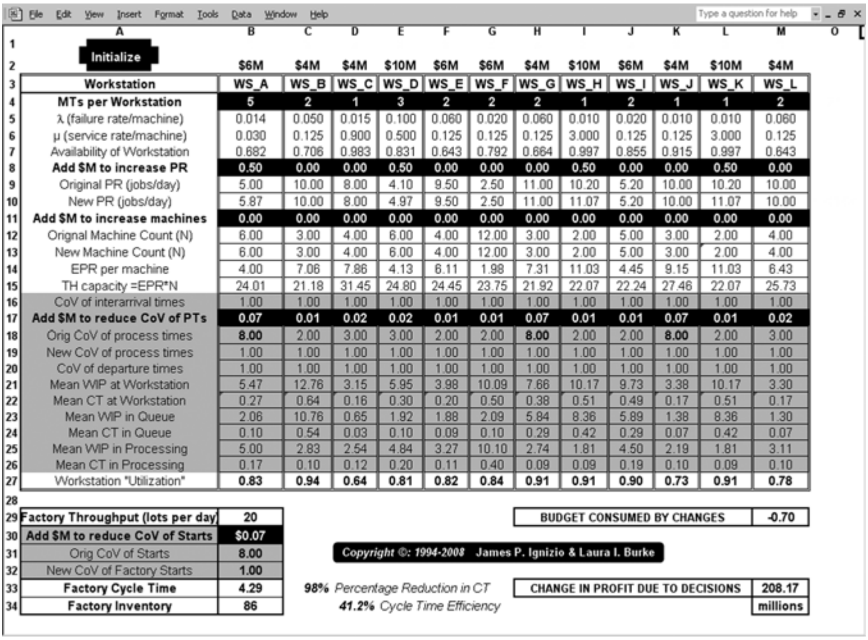
Considering the fact that there are invariably errors in collected data and that estimates of costs for projects (e.g., increasing raw process rates or reducing variability) and of the value of a day of cycle time are predictions, we probably should be satisfied with implementation of the heuristic approach to factory performance improvement. This is certainly evident in this example.

CHAPTER SUMMARY

The guidelines of Chapters 10 and 11, combined with the lessons from previous chapters, provide the basis for a practical, straightforward heuristic approach to factory performance improvement. Key to such improvement is, as demonstrated repeatedly, reducing variability (as well as any complexity serving to induce variability) in the factory.

FIGURE 12.8

A second heuristic solution.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Initialize												
2		$6M	$4M	$4M	$10M	$6M	$6M	$4M	$10M	$6M	$4M	$10M	$4M
3	**Workstation**	**WS_A**	**WS_B**	**WS_C**	**WS_D**	**WS_E**	**WS_F**	**WS_G**	**WS_H**	**WS_I**	**WS_J**	**WS_K**	**WS_L**
4	**MTs per Workstation**	**5**	**2**	**1**	**3**	**2**	**2**	**2**	**1**	**2**	**1**	**1**	**2**
5	λ (failure rate/machine)	0.014	0.050	0.015	0.100	0.060	0.020	0.060	0.010	0.020	0.010	0.010	0.060
6	μ (service rate/machine)	0.030	0.125	0.900	0.500	0.125	0.125	0.125	3.000	0.125	0.125	3.000	0.125
7	Availability of Workstation	0.682	0.706	0.983	0.831	0.643	0.792	0.664	0.997	0.855	0.915	0.997	0.643
8	**Add $M to increase PR**	**0.50**	**0.00**	**0.00**	**0.50**	**0.00**	**0.00**	**0.00**	**0.50**	**0.00**	**0.00**	**0.50**	**0.00**
9	Original PR (jobs/day)	5.00	10.00	8.00	4.10	9.50	2.50	11.00	10.20	5.20	10.00	10.20	10.00
10	New PR (jobs/day)	5.87	10.00	8.00	4.97	9.50	2.50	11.00	11.07	5.20	10.00	11.07	10.00
11	**Add $M to increase machines**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	Orignal Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
13	New Machine Count (N)	6.00	3.00	4.00	6.00	4.00	12.00	3.00	2.00	5.00	3.00	2.00	4.00
14	EPR per machine	4.00	7.06	7.86	4.13	6.11	1.98	7.31	11.03	4.45	9.15	11.03	6.43
15	TH capacity =EPR*N	24.01	21.18	31.45	24.80	24.45	23.75	21.92	22.07	22.24	27.46	22.07	25.73
16	CoV of interarrival times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	**Add $M to reduce CoV of PTs**	**0.07**	**0.01**	**0.02**	**0.02**	**0.01**	**0.01**	**0.07**	**0.01**	**0.01**	**0.07**	**0.01**	**0.02**
18	Orig CoV of process times	**8.00**	2.00	3.00	3.00	2.00	2.00	**8.00**	2.00	2.00	**8.00**	2.00	3.00
19	New CoV of process times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	CoV of departure times	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21	Mean WIP at Workstation	5.47	12.76	3.15	5.95	3.98	10.09	7.66	10.17	9.73	3.38	10.17	3.30
22	Mean CT at Workstation	0.27	0.64	0.16	0.30	0.20	0.50	0.38	0.51	0.49	0.17	0.51	0.17
23	Mean WIP in Queue	2.06	10.76	0.65	1.92	1.88	2.09	5.84	8.36	5.89	1.38	8.36	1.30
24	Mean CT in Queue	0.10	0.54	0.03	0.10	0.09	0.10	0.29	0.42	0.29	0.07	0.42	0.07
25	Mean WIP in Processing	5.00	2.83	2.54	4.84	3.27	10.10	2.74	1.81	4.50	2.19	1.81	3.11
26	Mean CT in Processing	0.17	0.10	0.12	0.20	0.11	0.40	0.09	0.09	0.19	0.10	0.09	0.10
27	Workstation "Utilization"	**0.83**	**0.94**	**0.64**	**0.81**	**0.82**	**0.84**	**0.91**	**0.91**	**0.90**	**0.73**	**0.91**	**0.78**
28													

	A	B
29	**Factory Throughput (lots per day)**	**20**
30	**Add $M to reduce CoV of Starts**	**$0.07**
31	Orig CoV of Starts	**8.00**
32	New CoV of Factory Starts	**1.00**
33	**Factory Cycle Time**	**4.29**
34	**Factory Inventory**	**86**

BUDGET CONSUMED BY CHANGES	**-0.70**
CHANGE IN PROFIT DUE TO DECISIONS	**208.17**
	millions

***Copyright ©: 1994-2008* James P. Ignizio & Laura I. Burke**

98% *Percentage Reduction in CT*
41.2% *Cycle Time Efficiency*

CASE STUDY 12: PAY THE PIPER

Tommy Jenkins receives his notice of termination by e-mail. It simply states that his services are no longer required and that, by 1 p.m. this day, he is to be escorted from the Factory 7 campus. As he gathers his belongings, limited to those that can fit in a single book box as per Muddle regulations, his office door swings open—without the courtesy of a knock.

"Are you ready, Tommy?" asks Ben Arnold, struggling to suppress a smile. "It's almost 1 p.m., and I've been asked to escort you from the premises."

Tommy Jenkins restrains his anger, tries hard to keep his feelings masked, and simply nods in the affirmative. There are no further words exchanged as Ben walks Tommy to the exit of the office complex.

As he turns the key in his plush Mercedes-Benz CL600, one word runs through Tommy's mind: *Unfair, unfair, unfair, unfair.*

Back in what had been Tommy's office, three words run through Ben's mind: *I did it! I did it! I did it!*

* * *

Donna Garcia is in a bad mood. Tommy Jenkins's removal had only been made possible by the coordinated efforts of her, Ben Arnold, and Jack Gibson. Their plan had worked perfectly, as far as the dismissal of Tommy had gone and his replacement by Ben. But one promise, a solemn promise made by both Ben and Jack, has not been honored.

They had promised her that once they were rid of Tommy, she would be promoted to the position of Ben's technical assistant. Instead, as she learns through a tersely worded e-mail, she is to be sent to Room 101 for reeducation. If that six-week course is completed successfully, she is to be reassigned to an aging Muddle factory site in, of all places, Fargo, North Dakota. There she will be responsible for decommissioning the facility. Her future after that has been left unspecified. There is, she thinks, no way I'm going to either Room 101 or Fargo.

Some 200 miles away, Jack Gibson, the junior factory manager for Factory 2, is informed that he has been promoted to the position of senior vice president and director of manufacturing, replacing William "Wild Bill" Barlow. This doesn't come as a surprise to Jack.

* * *

With the departure of Tommy Jenkins and the resignation of Donna Garcia, the leadership of the *LEAN* Forward team in Factory 7 is effectively gutted, at least in terms of its existence on the formal organization chart. This proves to be no problem because Marvin Muddle and the MRC conclude—*without any actual evidence*—that lean manufacturing has done absolutely nothing to improve the performance of their factories.

After much discussion, it is decided to disband the lean effort and to consider another approach—increased factory moves. The only, and most difficult, decision remaining is the choice of a new slogan and logo.

Marvin Muddle picks, from among the dozen or so slogans proposed, “Fast Mover.” The new logo will consist of the word *Muddle* located above a stylized drawing of a roadrunner. To most people, however, it just looks a lot like a confused and frightened bird.

An order is issued to all plant managers that, henceforth, the new measure of factory performance will be a count of moves per week through the machines in each factory. At the end of each quarter, the factory with the highest number of moves will be presented with the coveted “Muddle Badge of Merit” plus six-figure bonuses to the plant managers, along with funds sufficient to provide a free soft drink to each and every factory floor worker.

Ben Arnold is determined to win the award each and every quarter. He issues a memo to his department managers demanding that factory moves be increased by at least 20 percent. The department managers and factory floor personnel have already devised some rather ingenious methods for accomplishing, if not surpassing, those goals.

* * *

Julia Austen has called a meeting, and its participants have arrived at Winston Smith’s “war room.” She provides a brief summary of the situation now facing Factory 7, as well as the Muddle Corporation.

“Guys,” says Julia, “the good news is that Tommy Jenkins, Donna Garcia, and Bill Barlow are no longer with the firm. The bad news is that Ben Arnold is now our senior plant manager. But you already know all that. What you may not have heard is that Ben’s first order of business has been to demand that we increase factory moves by 20 percent or more. Naturally, he hasn’t explained just how this will be done. On the other hand, the good news is . . .”

“That’s insane,” Winston interrupts. “Focusing solely on factory moves, as we all know, is quite possibly the worst thing a factory can do. Besides, an increase in moves of 20 percent will overload this factory. If you think our performance is bad now, you ain’t seen nothin’ yet.”

“I’m sure we all agree,” says Julia. “But there is some good news, or at least I hope so. With all the fuss about the reorganizations, firings, and resignations, I think we can operate under the radar. Ben and his people are too busy with more important matters than to pay much attention to the three of us. We should be able to continue our work. At least that’s my opinion. One thing we might want to look into is the 20 percent increase in factory moves. I'd like to recommend we use Winston's models to investigate the impact of moves on the factory. Once that's done, why don't we just bite the bullet and present our findings to Ben and the other Factory 7 plant managers?"

"Frankly, I don't think we have anything to lose," says Dan. "We may be under the factory's radar at the moment, but sooner or later I'm sure the three of us will be looking for other jobs."

"I agree," says Winston. "We may as well give it a try. Who knows? Perhaps a miracle will happen. Perhaps we'll actually be able to convince management that we can vastly improve factory performance simply by reducing complexity and variability. That reminds me, I'm in the process of developing an executive summary of our findings. It's even better than I had thought."

"What's the bottom line?" asks Dan. "Just how much improvement did the simulation models indicate?"

"When we combine just four of Professor Leonidas' methods for reducing complexity and variability in the Factory 7 simulation model, we get a 72 percent reduction in factory cycle time and a 65 percent reduction in the variability of factory outs. In addition, mainly by the reduction in variability, we achieve a 9 percent increase in effective factory capacity. I'm also convinced that if we performed a Waddington analysis and brought our operating and maintenance specifications up to C^4U compliance, we could increase effective factory capacity by at least 20 percent—all this without having to buy any additional machines or hire any more personnel."

"Wow," says Julia, "that's amazing. "Fellows, we should put a slide-show presentation together. If these results don't convince plant management, nothing will."

CHAPTER 12 EXERCISES

Using the 12-workstation factory simulation model of this chapter, perform the following exercises.

1. Employ the greedy heuristic to find the maximal profit solution, but reverse the order of the first two phases (i.e., first reallocate maintenance personnel and then reduce variability).
2. Defend the use of a heuristic procedure (i.e., the greedy heuristic) as opposed to an optimization procedure.

This page intentionally left blank

CHAPTER 13

The Fundamental Model of Manufacturing

In preceding chapters I introduced, discussed, and illustrated the three fundamental equations of manufacturing. These equations, in turn, provided the foundation necessary for developing the pragmatic and cost-effective guidelines listed and illustrated in Chapters 10, 11, and 12. In this chapter I deal briefly with the fundamental model of manufacturing. Using this model, we shall be able to determine, precisely, the maximum theoretical capacity of a workstation (and thus a factory). By combining this model with an appreciation of variability, we will be better able to forecast the maximum sustainable capacity of a workstation or factory. Furthermore, by means of straightforward extensions to the fundamental model, it is possible to develop models that solve such problems as the optimization of operation-to-machine assignments (a.k.a. *dedications*).

While preceding chapters required only a limited appreciation of mathematics and employed only the most basic mathematical operations, readers should be forewarned that the fullest understanding of the material in this chapter is best achieved via at least a modest background in mathematical modeling and optimization [e.g., linear programming (Ignizio and Cavalier, 1994)]. Consequently, some readers may want to peruse the material, whereas others may have the background—and interest—sufficient to obtain a more complete appreciation of the fundamental model.

A REVIEW OF CAPACITY

The maximum theoretical capacity of a workstation was computed in preceding chapters, in certain instances, simply by adding the effective process rates (*EPRs*) of each machine. On another occasion, the harmonic mean was employed. In either case, the approach presented was limited to special cases. More specifically, we simply may sum the *EPRs* of the machines in a workstation to determine its maximum theoretical capacity if and only if

- The workstation supports only a single operation (i.e., process step).
- Each machine supports that single operation (even if at different process rates).

Table 13.1 presents data for a multiple-machine, single-operation (MM/SO) workstation in which the effective process rates may be added. Consequently, this workstation's *EPR* is $1.800 + 1.125 + 1.020 + 1.848 = 5.793$ lots per hour.

The harmonic mean, in turn, may be employed in instances in which

- The workstation supports multiple operations (i.e., process steps).
- Every machine in the workstation has the same process rate for a given operation (e.g., if machine A has a *PR* of 5 lots per hour on operation 4, then all other machines in the workstation must have a *PR* of 5 lots per hour on that particular operation).

Table 13.2 presents data for a multiple-machine, multiple-operation (MM/MO) workstation for which derivation of the capacity through employment of the harmonic mean is appropriate. This is

TABLE 13.1

Maximum Theoretical Capacity of an MM/SO Workstation

Machine	Process Rate (Lots/Hour)	Availability (Percent)	*EPR* (Lots/Hour) = *PR* • *A*
A	2	90	1.800
B	1.5	75	1.125
C	1.2	85	1.020
D	2.1	88	1.848

TABLE 13.2

Maximum Theoretical Capacity of an MM/MO Workstation

Machine	Operation 1 (Lots/Hour)	Operation 2 (Lots/Hour)	Operation 3 (Lots/Hour)	Harmonic Mean (*HM*)	Availability (*A*) (Percent)	EPR (Lots/Hour) = *HM* • *A*
A	2.0	1.2	1.8	1.5882	90	1.4294
B	2.0	1.2	1.8	1.5882	90	1.4294
C	2.0	1.2	1.8	1.5882	90	1.4294
D	2.0	1.2	1.8	1.5882	90	1.4294

so because the process rates for each operation (the numerical values in columns 2, 3, and 4) are the same for each machine. In other words, each machine is identical in terms of its process rate per operation type.

The maximum theoretical capacity of the workstation whose data are shown in Table 13.2 is found by summing the effective process rates of each machine (where these, in turn, must be determined by multiplying their harmonic means by their availability). Thus the maximum theoretical capacity of this workstation is 4 • 1.4294, or 5.7176 lots per hour.

THE FUNDAMENTAL MODEL

The fundamental model, or general capacity model (GCM), as developed by Ignizio (1992a, 1992b), may be employed in either of the two special cases described earlier or—more important—for the completely general case, that is, where

- The workstation supports one or more operations.
- Not every machine in the workstation necessarily has identical process rates for associated operations and even when some machines are not qualified (e.g., dedicated) to support certain operations.

Frankly, the GCM is so easy to formulate and solve (given access to a supporting linear programming software package such as SOLVER, which is a free add-on in the Microsoft Excel package) that the safest course of action is to always employ the GCM when computing workstation or overall factory maximum theoretical capacity. I shall initially illustrate application of GCM on the example originally provided in Table 13.2. While the harmonic mean happens to be appropriate for that workstation, it will be instructive to employ the GCM as an alternate (and preferred) approach.

For construction of the general form of the model, I shall impose a few initial assumptions, *each of which may be relaxed ultimately*. These initial assumptions are

- The average number of lots processed per hour through each operation is equal. (For example, if the workstation supports three operations and the average number of lots, requiring any of the three operations, is six lots per hour, then the average number of lots processed by each individual operation is two per hour.)
- Only one type of product is processed by the workstation.
- There is no scrap or loss owing to yield loss.

Under these initial assumptions, the maximum theoretical capacity of a workstation may be determined simply by maximizing the total number of lots per hour of any one of the operations supported by the workstation. Thus, if we let *OP1* represent the number of lots requiring operation 1 supportable by the workstation, our objective may be stated simply as "maximize *OP1*."

Other parameters used to construct the optimization model include

- i = machines ($i = 1, 2, \ldots, m$)
- j = operations ($j = 1, 2, \ldots, n$)
- $x(i,j)$ = number of hours devoted by machine i to the conduct of operation j
- $PR(i,j)$ = process rate in lots (or jobs or batches) per hour for the performance of operation j on machine i
- $OP(j)$ = total number of lots (or jobs or batches) requiring operation j that the workstation processes each hour

Employing these definitions, we may construct the form of the GCM for the situation depicted in Table 13.2. The resulting model is listed in Equations (13.1) through (13.10). As noted, these equations form a special type of optimization model known as a *linear program* (LP) (Ignizio and Cavalier, 1994).

$$\text{Maximize } OP1 \tag{13.1}$$

subject to

$$x(\text{A},1) + x(\text{A},2) + x(\text{A},3) \le 0.90 \tag{13.2}$$

$$x(\text{B},1) + x(\text{B},2) + x(\text{B},3) \le 0.90 \tag{13.3}$$

$$x(\mathrm{C},1) + x(\mathrm{C},2) + x(\mathrm{C},3) \le 0.90 \tag{13.4}$$

$$x(\mathrm{D},1) + x(\mathrm{D},2) + x(\mathrm{D},3) \le 0.90 \tag{13.5}$$

$$OP1 = PR(\mathrm{A},1) \bullet x(\mathrm{A},1) + PR(\mathrm{B},1) \bullet x(\mathrm{B},1) + PR(\mathrm{C},1) \bullet x(\mathrm{C},1) + PR(\mathrm{D},1) \bullet x(\mathrm{D},1) \tag{13.6}$$

$$OP2 = PR(\mathrm{A}, 2) \bullet x(\mathrm{A}, 2) + PR(\mathrm{B}, 2) \bullet x(\mathrm{B}, 2) + PR(\mathrm{C}, 2) \bullet x(\mathrm{C}, 2) + PR(\mathrm{D}, 2) \bullet x(\mathrm{D}, 2) \tag{13.7}$$

$$OP3 = PR(\mathrm{A}, 3) \bullet x(\mathrm{A},3) + PR(\mathrm{B},3) \bullet x(\mathrm{B},3) + PR(\mathrm{C},3) \bullet x(\mathrm{C},3) + PR(\mathrm{D},3) \bullet x(\mathrm{D},3) \tag{13.8}$$

$$OP1 = OP2 = OP3 \tag{13.9}$$

$$x(\mathrm{A},1), x(\mathrm{A},2), \ldots, x(\mathrm{D},2), x(\mathrm{D},3) \ge 0 \tag{13.10}$$

Each of the functions listed in the LP model may be defined as follows:

- Equation (13.1) is the LP objective function wherein we seek to find the maximum value of *OP1*, the number of lots supportable by the machines capable of processing operation 1. We could have just as well sought to maximize *OP2* or *OP3* because these variables are equal in value under our initial set of assumptions.
- Equations (13.2) through (13.5) state that the time devoted by a given machine (e.g., machine A in Equation 13.2) must be equal to or less than that available on the machine on average each hour. For example, in Equation (13.2), the time devoted by machine A to each of the three operations each hour must be equal to or less than the machine's average availability (i.e., 0.90 of an hour).
- Equations (13.6) through (13.8) serve to define the number of lots processed per hour for operations 1, 2, and 3, respectively.
- Equation (13.9) indicates that the number of lots processed per hour for each of the three operations must be equal (i.e., under the initial assumptions).
- Finally, Equation (13.10) says that the model's *decision variables* (i.e., hours allocated by each machine to each

operation) must be nonnegative. In other words, none of the hours allocated to an operation may have a negative value.

If we define one more parameter, we may derive the general form for this LP. We will let *TA*(*i*) represent the time available on machine *i* per hour for processing. This will be equal to the availability of the machine. The LP model for the MM/MO single-product workstation in general form then is: Find *x*(*i,j*) so as to maximize *OP1* (or *OP2, OP3,* etc.) such that

$$\sum_{j=1}^{n} x(i,j) \le TA(i) \qquad \text{for all } i \text{ machines}$$

$$OP(j) = \sum_{i=1}^{m} \left[PR(i,j) \bullet x(i,j) \right] \qquad \text{for all } j \text{ operations}$$

$$OP(s) = OP(t) \qquad \text{for all } s \text{ and } t$$

$$x(i,j) \ge 0 \qquad \text{for all } i \text{ and } j$$

SOLUTION VIA SOLVER

If you have access to an LP software package, the solution to the LP model (i.e., for the maximum theoretical capacity of a given workstation) may be found easily. To illustrate, examine how the model for Table 13.2 (i.e., Equations 13.1 through 13.10) is entered into an EXCEL spreadsheet and then solved by the SOLVER add-on. The associated spreadsheet problem representation is presented in Figure 13.1. Cell I20 contains the maximum theoretical capacity per operation, as determined by optimization, and has a value of 1.9059 lots per hour per operation. The maximum theoretical flow through all three operations is given in cell I21: 3 • 1.4059, or 5.7176 lots per hour per workstation.

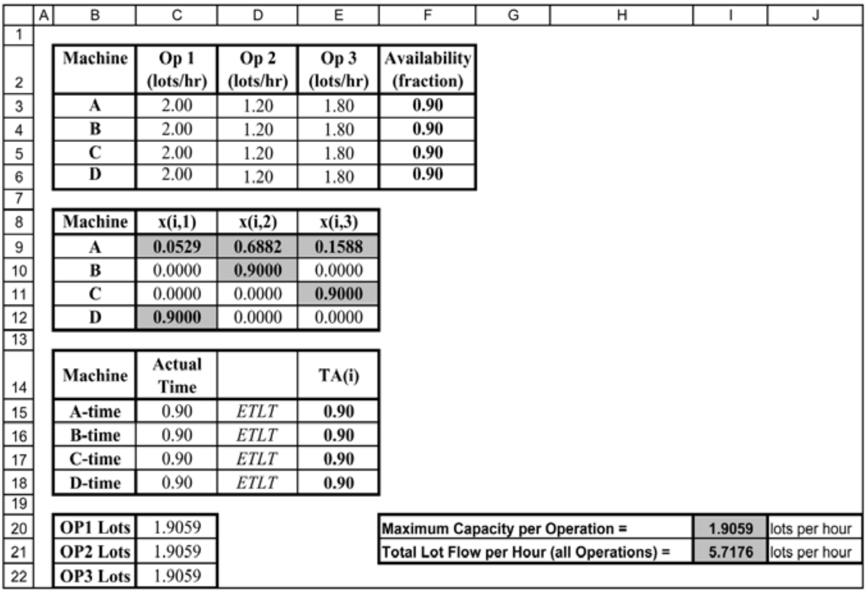
When we employed the harmonic mean for this situation, we computed the exact same maximum theoretical capacity for the entire workstation (i.e., 5.7176 lots per hour). Thus, at this point, the difference between the GCM and harmonic mean model is not apparent. I will deal with that matter later. First, however, examine Figure 13.2, which shows the formulas embedded in the spreadsheet.[1] The initial values of the decision variables (cells C9 through

---

[1] In Figures 13.1 and 13.2, the cells containing *ETLT* simply indicate that the values to the left should be equal to or less than (i.e., *ETLT*) those on the right. These *text entries* do not actually need to be included on the spreadsheet and have been inserted only as a reminder of the form of the corresponding constraints.

FIGURE 13.1

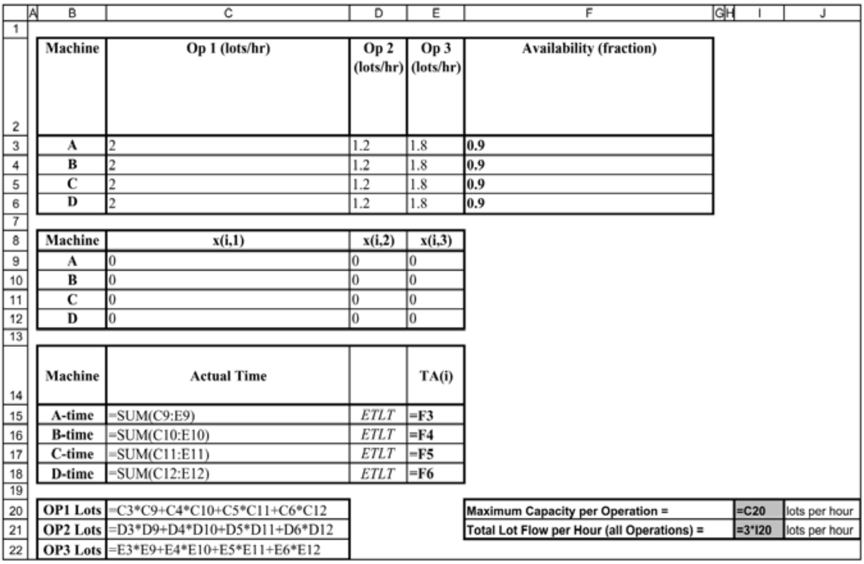
Spreadsheet representation of MM/MO capacity example.



	A	B	C	D	E	F	G	H	I	J
1										
2		**Machine**	**Op 1 (lots/hr)**	**Op 2 (lots/hr)**	**Op 3 (lots/hr)**	**Availability (fraction)**				
3		**A**	2.00	1.20	1.80	**0.90**				
4		**B**	2.00	1.20	1.80	**0.90**				
5		**C**	2.00	1.20	1.80	**0.90**				
6		**D**	2.00	1.20	1.80	**0.90**				
7										
8		**Machine**	**x(i,1)**	**x(i,2)**	**x(i,3)**					
9		**A**	**0.0529**	**0.6882**	**0.1588**					
10		**B**	0.0000	**0.9000**	0.0000					
11		**C**	0.0000	0.0000	**0.9000**					
12		**D**	**0.9000**	0.0000	0.0000					
13										
14		**Machine**	**Actual Time**		**TA(i)**					
15		**A-time**	0.90	*ETLT*	**0.90**					
16		**B-time**	0.90	*ETLT*	**0.90**					
17		**C-time**	0.90	*ETLT*	**0.90**					
18		**D-time**	0.90	*ETLT*	**0.90**					
19										
20		**OP1 Lots**	1.9059			**Maximum Capacity per Operation =**			**1.9059**	lots per hour
21		**OP2 Lots**	1.9059			**Total Lot Flow per Hour (all Operations) =**			**5.7176**	lots per hour
22		**OP3 Lots**	1.9059							

FIGURE 13.2

Spreadsheet formulas for example.



	A	B	C	D	E	F	G	H	I	J
1										
2		**Machine**	**Op 1 (lots/hr)**	**Op 2 (lots/hr)**	**Op 3 (lots/hr)**	**Availability (fraction)**				
3		**A**	2	1.2	1.8	**0.9**				
4		**B**	2	1.2	1.8	**0.9**				
5		**C**	2	1.2	1.8	**0.9**				
6		**D**	2	1.2	1.8	**0.9**				
7										
8		**Machine**	**x(i,1)**	**x(i,2)**	**x(i,3)**					
9		**A**	0	0	0					
10		**B**	0	0	0					
11		**C**	0	0	0					
12		**D**	0	0	0					
13										
14		**Machine**	**Actual Time**		**TA(i)**					
15		**A-time**	=SUM(C9:E9)	*ETLT*	**=F3**					
16		**B-time**	=SUM(C10:E10)	*ETLT*	**=F4**					
17		**C-time**	=SUM(C11:E11)	*ETLT*	**=F5**					
18		**D-time**	=SUM(C12:E12)	*ETLT*	**=F6**					
19										
20		**OP1 Lots**	=C3*C9+C4*C10+C5*C11+C6*C12			**Maximum Capacity per Operation =**			**=C20**	lots per hour
21		**OP2 Lots**	=D3*D9+D4*D10+D5*D11+D6*D12			**Total Lot Flow per Hour (all Operations) =**			**=3*I20**	lots per hour
22		**OP3 Lots**	=E3*E9+E4*E10+E5*E11+E6*E12							

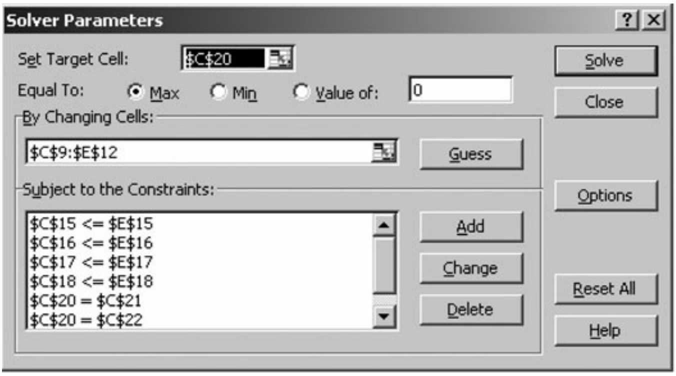
E12) have been set to zero in the spreadsheet. (Readers are advised to replicate this model and then employ SOLVER for its solution.)

The settings required in SOLVER are listed in Figures 13.3 and 13.4. (Not shown in Figure 13.3 is the final constraint: $C$21 = $C$22.)

There is one final important aspect of the solution found by the GCM that should be mentioned. Note in Figure 13.1 that the total number of nonzero-valued decision variables *x*(*i,j*) is six in the final solution. That portion of the spreadsheet is repeated in Table 13.3

FIGURE 13.3

SOLVER settings (objective, decision variables, and constraints).



Set Target Cell: $C$20

Equal To: ● Max ○ Min ○ Value of: 0

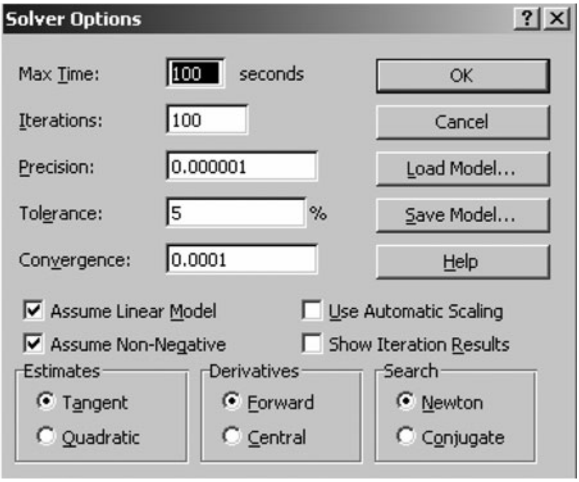
By Changing Cells: $C$9:$E$12

Subject to the Constraints:

$C$15 <= $E$15
$C$16 <= $E$16
$C$17 <= $E$17
$C$18 <= $E$18
$C$20 = $C$21
$C$20 = $C$22

FIGURE 13.4

SOLVER option settings



Max Time: 100 seconds
Iterations: 100
Precision: 0.000001
Tolerance: 5 %
Convergence: 0.0001

☑ Assume Linear Model
☐ Use Automatic Scaling
☑ Assume Non-Negative
☐ Show Iteration Results

Estimates: ● Tangent ○ Quadratic
Derivatives: ● Forward ○ Central
Search: ● Newton ○ Conjugate

TABLE 13.3

Decision Variables, Optimal Solution

Machine	$x(i,1)$	$x(i,2)$	$x(i,3)$
A	0.4742	0.0000	0.4258
B	0.0000	0.8000	0.0000
C	0.0000	0.1284	0.7516
D	0.9200	0.0000	0.0000

TABLE 13.4

Workstation Performance Data

Operations Machines	1 (Lots/Hour)	2 (Lots/Hour)	3 (Lots/Hour)	Availability (*A*) (Percent)
A	1.00	1.50	1.50	90
B	0.80	1.70	1.25	80
C	0.80	1.70	1.25	88
D	1.20	1.40	1.60	92

(where the variables taking on nonzero values are shown in the shaded cells).

The total number of nonzero variables, each representing the amount of time devoted by machine *i* to operation *j*, has an upper limit—in any optimal solution—to the total number of constraints in the LP model. Note that in Figure 13.3 (SOLVER settings) there are exactly seven constraints (recall that the seventh, \$C\$21=\$C\$22, is not displayed). Consequently, an optimal solution cannot have more than seven nonzero decision variables in the final solution. It will be shown later in the chapter that this is an important point in the development of solutions to such problems as optimally allocating jobs or operations to machines. First, however, let us examine a problem for which the employment of the harmonic means is *not* appropriate. Table 13.4 lists the performance data for a workstation consisting of four machines and supporting three operations.

Notice carefully that the machines do not, as was the case in Table 13.2, have identical process rates for their associated operations. For example, machine A has a process rate of 1 lot per hour for lots requiring operation 1, whereas machine B has a process rate of 0.8 lot per hour for that same operation. In this general case, you must employ the GCM if you hope to determine the correct value of the maximum theoretical capacity.

The optimization model for this problem is listed in Equations (13.11) through (13.20). Other than for the availability values and process rates, the model is identical to that solved previously.

$$\text{Maximize } OP1 \tag{13.11}$$

subject to

$$x(\text{A},1) + x(\text{A},2) + x(\text{A},3) \leq 0.90 \tag{13.12}$$

$$x(\text{B},1) + x(\text{B},2) + x(\text{B},3) \leq 0.80 \tag{13.13}$$

$$x(\text{C},1) + x(\text{C},2) + x(\text{C},3) \leq 0.88 \tag{13.14}$$

$$x(\text{D},1) + x(\text{D},2) + x(\text{D},3) \leq 0.92 \tag{13.15}$$

$$OP1 = PR(\text{A},1) \bullet x(\text{A},1) + PR(\text{B},1) \bullet x(\text{B},1) + PR(\text{C},1) \bullet x(\text{C},1) + PR(\text{D},1) \bullet x(\text{D},1) \tag{13.16}$$

$$OP2 = PR(\text{A},2) \bullet x(\text{A},2) + PR(\text{B},2) \bullet x(\text{B},2) + PR(\text{C},2) \bullet x(\text{C},2) + PR(\text{D},2) \bullet x(\text{D},2) \tag{13.17}$$

$$OP3 = PR(\text{A},3) \bullet x(\text{A},3) + PR(\text{B},3) \bullet x(\text{B},3) + PR(\text{C},3) \bullet x(\text{C},3) + PR(\text{D},3) \bullet x(\text{D},3) \tag{13.18}$$

$$OP1 = OP2 = OP3 \tag{13.19}$$

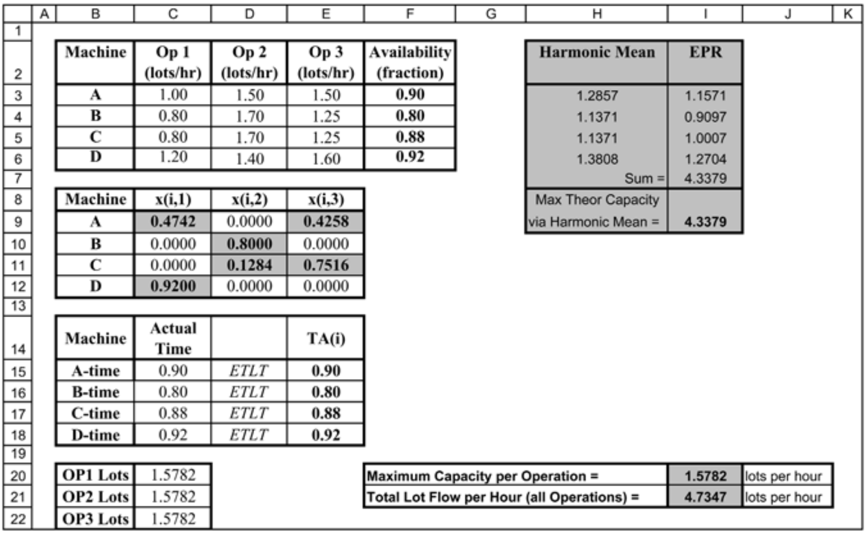
$$x(\text{A},1), x(\text{A},2), \ldots, x(\text{D},2), x(\text{D},3) \geq 0 \tag{13.20}$$

SOLUTION TO SECOND EXAMPLE

Employing the same approach as described previously, the model based on Table 13.4, that is, Equations (13.11) through (13.20), may be solved. The associated spreadsheet representation for the optimal solution is shown in Figure 13.5. Cell I20 contains the maximum theoretical capacity per operation, as determined by optimization, and has a value of 1.5782 lots per hour per operation. The maximum theoretical flow through all three operations is given in cell I21: 4.7347 lots per hour per workstation. The setup of the SOLVER menus happens to be, for this problem, identical to that employed in the preceding model.

FIGURE 13.5

Spreadsheet representation of MM/MO capacity example.



Machine	Op 1 (lots/hr)	Op 2 (lots/hr)	Op 3 (lots/hr)	Availability (fraction)
A	1.00	1.50	1.50	**0.90**
B	0.80	1.70	1.25	**0.80**
C	0.80	1.70	1.25	**0.88**
D	1.20	1.40	1.60	**0.92**

Machine	x(i,1)	x(i,2)	x(i,3)
A	**0.4742**	0.0000	**0.4258**
B	0.0000	**0.8000**	0.0000
C	0.0000	**0.1284**	**0.7516**
D	**0.9200**	0.0000	0.0000

Machine	Actual Time		TA(i)
A-time	0.90	*ETLT*	**0.90**
B-time	0.80	*ETLT*	**0.80**
C-time	0.88	*ETLT*	**0.88**
D-time	0.92	*ETLT*	**0.92**

OP1 Lots	1.5782
OP2 Lots	1.5782
OP3 Lots	1.5782

Harmonic Mean	EPR
1.2857	1.1571
1.1371	0.9097
1.1371	1.0007
1.3808	1.2704
Sum =	4.3379
Max Theor Capacity via Harmonic Mean =	**4.3379**

Maximum Capacity per Operation =	**1.5782**	lots per hour
Total Lot Flow per Hour (all Operations) =	**4.7347**	lots per hour

For purposes of comparison, the maximum theoretical capacity that would be computed if we erroneously (as indicated in the shaded cells in the top right-hand corner of the spreadsheet) employed the harmonic mean is given in cell I9 and is 4.3379 lots per hour through the workstation, or 1.4460 lots per hour per operation.

The correct maximum theoretical capacity of the workstation actually is 9 percent higher than that computed via the harmonic mean. In other words, if your firm employs the harmonic mean (or any capacity formula based on that mean), you would significantly underestimate the maximum theoretical capacity of this workstation.

MULTIPLE MACHINES, OPERATIONS, AND PRODUCTS

As promised, the GCM may be extended easily to encompass any type of workstation (e.g., multiple machines, multiple operations, multiple products, scrapped lots, inspection lots, etc.). This may be illustrated by means of modeling a problem involving multiple machines, multiple operations, and multiple products. The data for the problem used to illustrate the procedure are listed in Table 13.5.

The problem now faced involves two products, *X* and *Y*. We assume that product *X* must comprise 80 percent of the total volume supported by the workstation, whereas product *Y* forms the

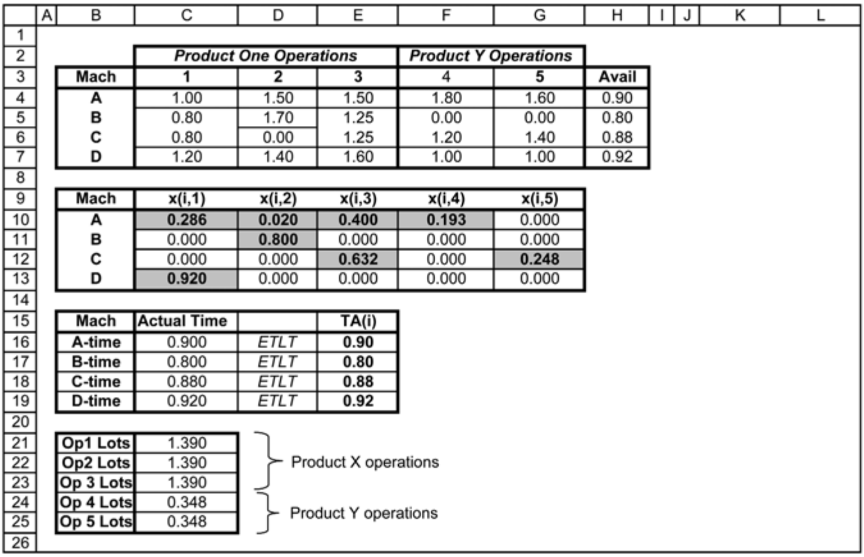
TABLE 13.5

Multiple Machines, Operations, and Products

Operations / Machines	Product *X*			Product *Y*		
	1 (Lots/Hour)	2 (Lots/Hour)	3 (Lots/Hour)	4 (Lots/Hour)	5 (Lots/Hour)	Availability (*A*) (Percent)
A	1.00	1.50	1.50	1.80	1.60	90
B	0.80	1.70	1.25	0.00	0.00	80
C	0.80	0.00	1.25	1.20	1.40	88
D	1.20	1.40	1.60	1.00	1.00	92

FIGURE 13.6

Optimal solution to multiproduct example.



	Product One Operations			*Product Y Operations*		
Mach	1	2	3	4	5	Avail
A	1.00	1.50	1.50	1.80	1.60	0.90
B	0.80	1.70	1.25	0.00	0.00	0.80
C	0.80	0.00	1.25	1.20	1.40	0.88
D	1.20	1.40	1.60	1.00	1.00	0.92

Mach	x(i,1)	x(i,2)	x(i,3)	x(i,4)	x(i,5)
A	0.286	0.020	0.400	0.193	0.000
B	0.000	0.800	0.000	0.000	0.000
C	0.000	0.000	0.632	0.000	0.248
D	0.920	0.000	0.000	0.000	0.000

Mach	Actual Time		TA(i)
A-time	0.900	*ETLT*	0.90
B-time	0.800	*ETLT*	0.80
C-time	0.880	*ETLT*	0.88
D-time	0.920	*ETLT*	0.92

Op1 Lots	1.390
Op2 Lots	1.390
Op 3 Lots	1.390
Op 4 Lots	0.348
Op 5 Lots	0.348

remaining 20 percent. Product *X* requires three operations (1, 2, and 3), whereas product *Y* involves two (4 and 5). While machines A, C, and D may be used to support the operations of either product, machine B may be used only in support of product *X* (note the shaded cells for machine B and operations 4 and 5). Finally, machine C is unable to support operation 2 of product *X* (indicated by the shaded cell under operation 2).

The optimal solution—the maximum theoretical capacity of this workstation—is shown in Figure 13.6. Given the condition that products *X* and *Y* comprise, respectively, 80 and 20 percent of the volume of product types supported by the workstation, the maximum theoretical capacity for each of the three operations associated with product X is 1.390 lots per hour, whereas that for each of the two operations for product Y is 0.348 lots per hour. Multiplying each of these results by the number of operations per product, we find that the maximum theoretical capacity of the entire workstation is, for the given product mix, 4.866 lots per hour.

The optimization model for this multiproduct example is shown in Equations (13.21) through (13.36). These equations follow the same pattern as employed in the solution of the MM/MO single-product workstation. Particular attention, however, should be paid to Equations (13.31) through (13.35). These serve to maintain the 80–20 proportions of the two products.

$$\text{Maximize } OP1 \tag{13.21}$$

subject to

$$x(\mathrm{A},1) + x(\mathrm{A},2) + x(\mathrm{A},3) + x(\mathrm{A},4) + x(\mathrm{A},5) \le 0.90 \tag{13.22}$$

$$x(\mathrm{B},1) + x(\mathrm{B},2) + x(\mathrm{B},3) + x(\mathrm{B},4) + x(\mathrm{B},5) \le 0.80 \tag{13.23}$$

$$x(\mathrm{C},1) + x(\mathrm{C},2) + x(\mathrm{C},3) + x(\mathrm{C},4) + x(\mathrm{C},5) \le 0.88 \tag{13.24}$$

$$x(\mathrm{D},1) + x(\mathrm{D},2) + x(\mathrm{D},3) + x(\mathrm{D},4) + x(\mathrm{D},5) \le 0.92 \tag{13.25}$$

$$OP1 = PR(\mathrm{A},1) \bullet x(\mathrm{A},1) + PR(\mathrm{B},1) \bullet x(\mathrm{B},1) + PR(\mathrm{C},1) \bullet x(\mathrm{C},1) + PR(\mathrm{D},1) \bullet x(\mathrm{D},1) \tag{13.26}$$

$$OP2 = PR(\mathrm{A},2) \bullet x(\mathrm{A},2) + PR(\mathrm{B},2) \bullet x(\mathrm{B},2) + PR(\mathrm{C},2) \bullet x(\mathrm{C},2) + PR(\mathrm{D},2) \bullet x(\mathrm{D},2) \tag{13.27}$$

$$OP3 = PR(\mathrm{A},3) \bullet x(\mathrm{A},3) + PR(\mathrm{B},3) \bullet x(\mathrm{B},3) + PR(\mathrm{C},3) \bullet x(\mathrm{C},3) + PR(\mathrm{D},3) \bullet x(\mathrm{D},3) \tag{13.28}$$

$$OP4 = PR(\mathrm{A},4) \bullet x(\mathrm{A},4) + PR(\mathrm{B},4) \bullet x(\mathrm{B},4) + PR(\mathrm{C},4) \bullet x(\mathrm{C},4) + PR(\mathrm{D},4) \bullet x(\mathrm{D},4) \tag{13.29}$$

$$OP5 = PR(\mathrm{A},5) \bullet x(\mathrm{A},5) + PR(\mathrm{B},5) \bullet x(\mathrm{B},5) + PR(\mathrm{C},5) \bullet x(\mathrm{C},5) + PR(\mathrm{D},5) \bullet x(\mathrm{D},5) \tag{13.30}$$

$$OP1 = OP2 = OP3 \tag{13.31}$$

$$OP1 = 4 \bullet OP4 = 4 \bullet OP5 \tag{13.32}$$

$$OP2 = 4 \bullet OP4 = 4 \bullet OP5 \tag{13.33}$$

$$OP3 = 4 \bullet OP4 = 4 \bullet OP5 \tag{13.34}$$

$$OP4 = OP5 \tag{13.35}$$

$$x(\mathrm{A},1), x(\mathrm{A},2), \ldots, x(\mathrm{D},4), x(\mathrm{D},5) \geq 0 \tag{13.36}$$

Readers may wish to enter the equivalent mathematical model into SOLVER and employ that package to determine the solution—and compare those results with the results presented in Figure 13.6. The modeling process and SOLVER menus follow a format similar to those employed in the two preceding examples. The problem, in this case, is simply a bit larger in terms of the number of variables and constraints. It should be noted, however, that solutions to models larger than those dealt with thus far may require a different version of SOLVER (e.g., Premium SOLVER) or other LP software capable of dealing with more variables and constraints.

THE FALLACY OF FLEXIBILITY

If we examine the optimal solutions depicted in either Figures 13.1, 13.5, or 13.6, one thing should stand out. This is the fact that the capacity of a workstation is maximized without the need to have every machine dedicated to the support of every operation. This conclusion may fly in the face of conventional wisdom, wherein it may be believed that for the sake of both capacity and *flexibility,* every machine must be capable of supporting every operation assigned to a workstation if its capacity is to be maximized.

The results indicated in Figures 13.1, 13.5, and 13.6 demonstrate that this is not the case; that is, the maximum theoretical workstation capacity was achieved even though some machines were not assigned to the support of some operations. It may be (and has been) demonstrated by means of factory simulations, in fact, that overall factory performance may be improved—sometimes significantly—through the employment of workstation operation-to-machine allocations in which not every machine supports every operation.

These optimal operation-to-machine allocations may be achieved by means of extensions of LP-based models (the GCM model) discussed previously. In this case, however, the models are no longer strictly linear and require software capable of dealing with integer and, in some cases, even nonlinear mathematical models.

OPERATION-TO-MACHINE DEDICATIONS: AN OVERVIEW

In a factory employing reentrancy, determining the allocation of operations to machines (e.g., machine dedications or qualifications) may play a large role in determining workstation and factory performance. In many firms, this decision would appear to be accomplished by heuristic means—most of which seem to have little basis other than judgment and luck. There is a better way. First, however, consider the inherent complexity of the operation-to-machine dedication problem.

Examine, for example, an extremely simple workstation consisting of two machines and two operations. Given no other restrictions, the 16 possible operation-to-machine dedication schemes are listed in Table 13.6. For example, in the first operation-to-machine

TABLE 13.6

Allocation Schemes, Two Machines and Two Operations

Operation-Machine Scheme	Machine 1	Machine 2
1	None	1,2
2	None	1
3	None	2
4	None	None
5	1	None
6	1	1
7	1	2
8	1	1, 2
9	2	None
10	2	1
11	2	2
12	2	1, 2
13	1, 2	None
14	1, 2	1
15	1, 2	2
16	1, 2	1, 2

allocation, machine 1 supports no operations, whereas machine 2 supports two. The fourteenth scheme has machine 1 supporting both operations, whereas machine 2 supports only operation 1.

While not all these schemes are feasible (i.e., schemes 2, 3, 4, 5, 6, 9, and 11 do not provide support for all operations), the table begins to indicate the fact that the operation-to-machine allocation problem is not nearly so simple as it might first appear. More specifically, the operation-to-machine allocation problem is a type of problem known to be *NP-complete,* or in other words, it is *combinatorially explosive.* In plain English, this means that for almost any real-world situation (e.g., a highly reentrant factory), this is an extremely difficult problem to solve.

In general, given M machines and O operations, the total number of operation-to-machine allocation schemes is given as

$$\text{Number of schemes} = 2^{M \bullet O}$$

Thus, in a somewhat more realistic workstation consisting of, say, 16 machines and 12 operations, the number of allocation schemes would be

Number of Boolean (0/1) variables $= M \bullet O = 12 \bullet 16 = 192$
Number of operation-to-machine schemes $= 2^{192}$
or roughly 6.28e+57.

Consequently, even if we employed one of the world's fastest computers (e.g., running at 35,600 gigaflops per second, i.e., 35,600 billion floating-point operations per second), and even if that supercomputer could evaluate each allocation scheme via only a single floating-point operation (in reality, it would take far more operations), it still would take more than $5.6 \bullet 10^{36}$ years to evaluate all possible allocations. Clearly, the determination of operation-to-machine allocation schemes is not a trivial problem.

The number of constraints that must be employed to describe the problem fully further increases its complexity. For example, any schemes that do not provide support for all operations must be ruled out immediately. In addition there may be operation-to-operation conditions or conflicts. For example, operations X and Y always may have to be performed on the same machine, whereas perhaps operations A and B never can be performed on the same machine.

Another possible constraint may be a result of the physical limitations of the machines in the workstation. For example, there may be a limit on the number of chemicals for which a machine is plumbed. This situation occurs, for example, with photolithography machines, in which no more plumbing than for, say, four different photoresists may be accommodated. Consequently, any operation-to-machine allocation scheme that requires more than four photoresists for that machine would be infeasible.

There are also practical limitations to how many operations should be supported by a given machine and the fact that it takes time and personnel to perform and maintain the operation-to-machine qualifications (e.g., periodic tests, calibrations, etc.) within the respective process specifications. Taken together, the number of possible schemes coupled with practical limitations may serve to form a massively large problem of combinatorial optimization.

Simply put, there is little likelihood that any heuristic approach to the allocation of operations to machines will come close to optimal. In fact, it is almost certain that heuristically derived schemes will be far less effective than the optimal allocation. As such, it would seem reasonable to find an effective, practical way in which to optimally allocate operations to machines.

The following section presents the general mathematical model describing the operation-to-machine problem. While it is based on the GCM of preceding sections, the problem is considerably more complex—as is the mathematical model. In addition, we are no longer able to employ the simple LP model and associated software cited previously. There are, however, a number of commercially available software packages that may be employed to derive optimal or near-optimal solutions. I now proceed to a description of the model for a fairly typical operation-to-machine allocation problem.

OPERATION-TO-MACHINE DEDICATIONS: THE BASIC MODEL

It should be noted that the general form of mathematical model to be presented is not new. The model was developed originally for the representation and solution of reentrant networks (such as certain business processes and supply chains) in the 1990s (Ignizio, 1992b).

While a reentrant business process or supply chain and a reentrant factory would seem to be quite different, their mathematical representations are similar (i.e., either may be represented by a network that includes feedback/reentrant loops). As a consequence, the same fundamental Boolean optimization model (a model in which the variables may take on values of only 0 or 1) employed for supply chains may be employed for the operation-to-machine allocation model.

A quite basic form of the Boolean optimization model that serves to define the operation-to-machine allocation problem is presented below. It is assumed that the problem involves jobs that arrive in the form of lots and that the time period of interest is a 168-hour week. I begin, however, with definitions of the parameters that serve to support the model.

Definitions

- $r(i,j,k)$ = a Boolean variable where $r(i,j,k)$ is 1 if operation k is performed on machine i of lot j during the week and is 0 otherwise.
- $y(i,k)$ = a Boolean variable where $y(i,k)$ is 1 if operation k is qualified on machine i and is 0 otherwise.
- $x(i,r)$ = a Boolean variable where $x(i,r)$ is 1 if machine i uses a chemical (e.g., photoresist) r and is 0 otherwise.
- $\theta(r)$ = the set of operations that require a chemical of type r.
- $R_{\max}(i)$ = the maximum number of chemicals of any type that may be allocated to machine i.
- $a(i,k)$ = time required (in hours) for performance of operation k per lot on machine i. Note that this includes any additional average time for rework, test lots, and setups.
- $TA(i)$ = time available per week for performance of any and all operations assigned to machine i. (*Note:* Assume this is 168 hours per week times the availability of the machine.)
- $T(k)$ = the minimum amount of time that must be made available each week for the conduct of operation k, including additional time for rework at the operation (this, in turn, is a function of the desired throughput of the factory).
- *gap* = minimum gap across all the machines (note that the gap is defined as the difference between the time available on the machine and the time consumed by the operations performed by the machine).
- λ = a small multiplier (e.g., 0.0001 in our case) used to control the values of $x(i,r)$, that is, used in the support of the transformation of a nonlinear function into a linear function.
- M = a large multiplier (e.g., 1,000 in our case) used in support of the transformation of a nonlinear function into a linear function.

- And where
    - $k = 1, \ldots, K\ i = 1, \ldots, m\ j = 1, \ldots, n\ r = 1, \ldots, R$

With these definitions behind us, the formulation of the model to be employed may be presented:

$$\text{Maximize } [gap - \lambda \bullet x(i,r)] \tag{13.37}$$

subject to

$$\sum_{k=1}^{K}\sum_{j=1}^{n}\left[r(i,j,k)\bullet a(i,k)\right] + gap \le TA(i) \quad \text{for all } i \tag{13.38}$$

$$y(i,k)\bullet M \ge \sum_{j=1}^{n} r(i,j,k) \quad \text{for all } i \text{ and } k \tag{13.39}$$

$$\sum_{i=1}^{m} y(i,k) \ge 2 \quad \text{for all } k \tag{13.40}$$

$$\sum_{k=1}^{K} y(i,k) \le \begin{array}{l}\text{maximum number of operations on machine } i\\ \text{for all } i\end{array} \tag{13.41}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left[r(i,j,k)\bullet a(i,k)\right] \ge T(k) \quad \text{for all } k \tag{13.42}$$

$$\sum_{i=1}^{m} r(i,j,k) = 1 \quad \text{for all } j \text{ and } k \tag{13.43}$$

$$\sum_{k=\theta(r)} y(i,k) \le x(i,r)\bullet M \quad \text{for all } j \text{ and } r \tag{13.44}$$

$$\sum_{r=1}^{R} x(i,r) \le R_{\max}(i) \tag{13.45}$$

where $x(i,r)$ and $y(i,k)$ are 0–1 (i.e., Boolean) variables.

We also may add numerous other conditions such as those one would encounter in a real-world factory. For example:

- $y(i,5) + y(i,7) \le 1$; that is, operations 5 and 7 cannot both be performed on machine *i*.
- $y(3,3) - y(3,9) = 0$; that is, if operation 3 is performed on machine 3, then operation 9 also must be performed on machine 3.

Each of the functions in the model is described briefly in Table 13.7, and this is followed by a numerical illustration.

TABLE 13.7

Model Components Definitions

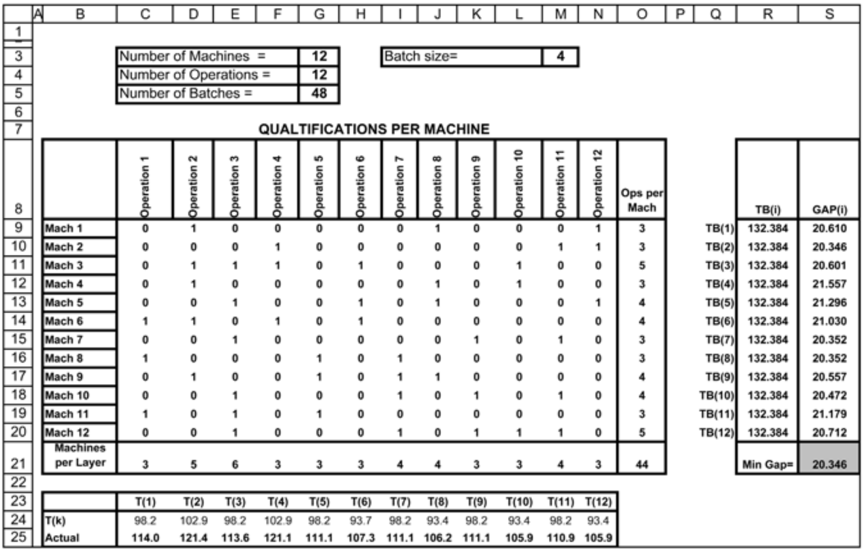
Function	Description
(13.37)	Objective function: We seek to maximize the minimum gap across the workstation; that is, balance the workload across the set of machines to minimize factory variability. Subtracted from this is the number of chemicals across the workstation multiplied by some small number (e.g., to set the number of chemicals per machine to zero unless absolutely required to support the constraint set).
(13.38)	Constraint: Limit the time devoted to all the operations on a given machine for the week to less than the total time available on that machine.
(13.39)	Constraint: Ensures that a dedication is made to a machine only if necessary.
(13.40)	Constraint: Ensures that at least two machines are qualified for every operation to maintain redundancy (this number may be adjusted as desired).
(13.41)	Constraint: Limits the maximum number of dedications on each machine.
(13.42)	Constraint: Requires that the time devoted to the operations on the lots equals or exceeds the minimum time required for that week.
(13.43)	Constraint: Ensures that every operation of every machine is supported.
(13.44)	Constraint: Ensures that the variable $x(i,r)$ is set to a value of 1 if and only if this is necessary to satisfy other constraints.
(13.45)	Constraint: The total number of chemicals employed by each machine must be less than its maximum plumbed capacity.
Others	Other constraints (e.g., limitations on the minimum or maximum number of machines qualified per operation or operations qualified per machine).

OPERATION-TO-MACHINE DEDICATIONS: AN ILLUSTRATION

Figure 13.7 presents the final solution for a moderately sized operation-to-machine allocation problem. There are 12 machines in a workstation that happens to support 12 operations. Jobs arrive in batches, with each batch consisting of 4 jobs. Given the process rates and other supporting data (not shown), a 0 in the qualification matrix indicates that the associated machine is not to be qualified for the given operation (e.g., machine 1 will not support operation 1), whereas a 1 means that there is an operation-to-machine allocation (e.g., machine 1 should be qualified to support operation 2).

FIGURE 13.7

Optimal operation-to-machine dedications,



Number of Machines = 12
Number of Operations = 12
Number of Batches = 48
Batch size= 4

QUALTIFICATIONS PER MACHINE

		Operation 1	Operation 2	Operation 3	Operation 4	Operation 5	Operation 6	Operation 7	Operation 8	Operation 9	Operation 10	Operation 11	Operation 12	Ops per Mach		TB(i)	GAP(i)
9	Mach 1	0	1	0	0	0	0	0	1	0	0	0	1	3	TB(1)	132.384	20.610
10	Mach 2	0	0	0	1	0	0	0	0	0	0	1	1	3	TB(2)	132.384	20.346
11	Mach 3	0	1	1	1	0	1	0	0	0	1	0	0	5	TB(3)	132.384	20.601
12	Mach 4	0	1	0	0	0	0	0	1	0	1	0	0	3	TB(4)	132.384	21.557
13	Mach 5	0	0	1	0	0	1	0	1	0	0	0	1	4	TB(5)	132.384	21.296
14	Mach 6	1	1	0	1	0	1	0	0	0	0	0	0	4	TB(6)	132.384	21.030
15	Mach 7	0	0	1	0	0	0	0	0	1	0	1	0	3	TB(7)	132.384	20.352
16	Mach 8	1	0	0	0	1	0	1	0	0	0	0	0	3	TB(8)	132.384	20.352
17	Mach 9	0	1	0	0	1	0	1	1	0	0	0	0	4	TB(9)	132.384	20.557
18	Mach 10	0	0	1	0	0	0	1	0	1	0	1	0	4	TB(10)	132.384	20.472
19	Mach 11	1	0	1	0	1	0	0	0	0	0	0	0	3	TB(11)	132.384	21.179
20	Mach 12	0	0	1	0	0	0	1	0	1	1	1	0	5	TB(12)	132.384	20.712
21	Machines per Layer	3	5	6	3	3	3	4	4	3	3	4	3	44		Min Gap=	20.346

		T(1)	T(2)	T(3)	T(4)	T(5)	T(6)	T(7)	T(8)	T(9)	T(10)	T(11)	T(12)
24	T(k)	98.2	102.9	98.2	102.9	98.2	93.7	98.2	93.4	98.2	93.4	98.2	93.4
25	Actual	114.0	121.4	113.6	121.1	111.1	107.3	111.1	106.2	111.1	105.9	110.9	105.9

Note that only 44 of the possible 144 qualifications (i.e., all 12 machines qualified for all 12 operations) are employed in the optimal solution.

Based on extensive simulations, the employment of optimal operation-to-machine dedications in place of more common approaches (e.g., intuition, guesses, etc.) indicates that factory performance is often improved anywhere from 5 to 20 percent through the optimization of operation-to-machine dedications. The reason for this lies primarily in the balance in the loading of the individual machines in the workstation.

This approach to balance should not be confused with the notion of balanced production lines, however (i.e., the so-called fundamental premise of lean manufacturing). Here we are striving to produce equal gaps (differences between the time available on individual machines and the time consumed in processing). Balanced production lines, on the other hand, seek to have equal cycle times among all the workstations in the production line.

To conclude our discussion, I now indicate how to estimate the maximum sustainable capacity of a workstation. This may be done once the maximum theoretical capacity has been determined, as has been described.

ESTIMATING MAXIMUM SUSTAINABLE CAPACITY

The GCM provides us with an accurate determination of the maximum theoretical capacity of a workstation and, by extension, that of the factory. By combining it with the three fundamental equations of manufacturing, we may—at least in theory—estimate the maximum sustainable capacity.

Consider, for example, the most simple factory type, one in which there is no batching, no reentrancy, and only one machine supports a given process step. The formula for the cycle time of such a process step is given by the most basic form of the second fundamental equation of manufacturing:

$$CT_{ps} = \underbrace{\left(\frac{C_{AR}^2 + C_{EPT}^2}{2}\right) \bullet \left[\frac{\rho}{(1-\rho)}\right] \bullet \left(\frac{1}{EPR_{ps}}\right)}_{\text{wait in queue time}} + \underbrace{\left(\frac{1}{EPR_{ps}}\right)}_{\text{effective process time}} \tag{13.46}$$

Recall from Chapter 3 that the occupancy rate (i.e., utilization) of these machines is expressed by

$$\rho = \frac{AR}{EPR} \tag{13.47}$$

By using Equations (13.46) (and assuming that arrival rate and effective process-time variability are known) and (13.47), it may be shown that the maximum permissible value of the arrival rate for the maximum tolerable cycle time of the process step is

$$AR_{\max} \le EPR \bullet \rho_{\max} \tag{13.48}$$

The value for $\rho_{\max}$ is found by substitution into Equation (13.46) (given the maximum permissible value for the cycle time of the process step). The value of *EPR* is found by means of the GCM. Thus, assuming that $\rho_{\max}$ is 90 percent and the value of *EPR* is 3 lots per hour, the maximum permissible arrival rate is

$$AR_{\max} \le EPR \bullet \rho_{\max} = 0.90 \bullet 3 \text{ lots/hour} = 2.7 \text{ lots/hour}$$

In other words, the maximum sustainable capacity at this process step is 2.7 lots per hour.

To determine the maximum sustainable capacity of the entire production line would require the summing of the formulas of cycle times for each of the process steps—given the maximum permissible value of total factory cycle time. For a simple factory with relatively few process steps, this could, in practice, be computed. More realistic factories with many process steps would be more involved, but the solution still would be possible theoretically.

In the real world, the time and effort involved most likely would not be worthwhile or even necessary. More specifically, virtually all the terms in the fundamental equations are estimates—some of which may be very rough estimates. As such, rather than attempting to determine the maximum sustainable capacity of a workstation precisely, it is, I believe, more rational to employ an approximation.

Table 13.8 lists approximations that have produced more than adequate results for certain classes of real-world factories (and decent approximations for most all types). The values for the coefficients of variability (*CoV*) for the last three columns are found by finding the average of the *CoV* of job arrivals at a processing entity (e.g., a workstation) and that of the effective process times of the machines forming that entity.[2]

To illustrate, assume that a workstation in the factory of interest has a maximum theoretical capacity (found by means of the GCM) of 4,000 units per week. Furthermore, the average coefficient of variability of job arrivals at the workstation is 1.4, whereas the average coefficient of variability of its effective process times is 2.2. The average of these two values (i.e., $2.2 + 1.4 \div 2$) is 1.8. This means that the factor that should be used is, from Table 13.8, a value of 0.90. To estimate the maximum sustainable capacity, we multiply the maximum theoretical capacity by this factor. The

TABLE 13.8

Maximum Sustainable Capacity Factors

	$CoV < 1$	$1 \le CoV < 2$	$2 \le CoV < 3$
Factor	0.93	0.90	0.87

---

[2] Alternately, you may use the average of the *CoV* values of job arrivals (e.g., at a factory) and job departures. This also seems to produce reasonable estimates.

result (4,000 jobs per week × 0.9) is 3,600 jobs per week and represents a reasonable estimate of the maximum sustainable capacity of the workstation. The maximum sustainable capacity of the factory, in turn, is simply the minimum of those of all the workstations in the factory.

While numerous other extensions of the GCM may be developed, I conclude the discussion of this topic here.

CHAPTER SUMMARY

The GCM may be used to determine the maximum theoretical capacity of a given workstation and thus the maximum theoretical capacity of an entire factory. Extensions of the GCM permit the modeling and solution of a number of important factory problems, including a determination of the optimal operation-to-machine qualification problem. Finally, by means of an adjustment factor, one may develop reasonably accurate estimates of the maximum sustainable capacity of a workstation or an entire factory.

CASE STUDY 13: IT’S SHOWTIME

Ben Arnold and his two fellow plant managers are uncharacteristically quiet during Winston’s presentation. Not only does Winston claim that factory performance may be improved immensely by means of a few straightforward and inexpensive efforts, but he also has the audacity to state that an increase in factory moves—by itself, as ordered by Ben—will absolutely destroy performance. Finishing the presentation, Winston asks if there are any questions.

Ben, smirking, responds, “Are you three out of your mind? Just because those weird ideas worked on your simulation models means absolutely nothing. None of you has ever run a real factory, just those silly simulation models. My intuition, and it never fails me, is that your nutty proposal likely will destroy us.”

“But,” argues Dan, “Winston has just showed you that increasing factory moves definitely will destroy this factory. Don’t you remember what happened when Tommy Jenkins increased factory starts? What will it take to convince you? Do you really want to help this company?”

“Mr. Ryan,” says Ben, “you’re fired. I want you off this campus by the end of the workday.”

Winston and Julia realize that there is no point in any further discussion. Each is convinced that he or she will be the next to be fired.

"As for you two," says Ben, pointing to Julia and Winston, "you've used unauthorized simulation models in support of your efforts. That's a clear violation of this firm's 'No Deviations' policy. You've also rebuilt company computers, computers that could have been sold for scrap. That's likely a felony, although I'll have to check with our corporate lawyers."

"Are we all fired?" asks Julia.

"For the moment, only Mr. Ryan is terminated," Ben replies. "My colleagues and I will discuss the matter. You'll have our decision by the end of the week."

* * *

Julia and Winston are in the "war room." The silence is broken by a question from Julia.

"Winston, why didn't they just fire us on the spot, like they did with Dan?"

"My guess is that they may be concerned that we might hire on to one of Muddle's competitors. I'm guessing that's what they're discussing right now."

"Winston, dear Winston, why don't *we* make this decision? Why wait for them? Why on earth should we stay with this company? As long as it's run by people like Marvin Muddle, Jack Gibson, and Ben Arnold, nothing is going to change."

CHAPTER 13 EXERCISES

1. Employ LP (e.g., the SOLVER package) to check and verify the results obtained for the maximum theoretical capacities of the workstations depicted in Tables 13.2, 13.4, and 13.5.
2. A factory has a coefficient of variability for factory starts of 3.0 and an average coefficient of variability for effective process times of 2.3. The maximum theoretical capacity of the factory's constraint is 10,000 units per week. What is its estimated maximum sustainable capacity?

This page intentionally left blank

CHAPTER 14

The Elements of Success

As has been emphasized repeatedly, the three enemies of factory performance are variability, complexity, and lackluster leadership. The preceding chapters have dealt mostly with ways in which to identify and mitigate variability and complexity. Less coverage has been devoted, however, to dealing with lackluster leadership or, in particular, the specific details of how to conduct an efficient, successful, and sustainable factory performance-improvement effort. I attempt to rectify these omissions in this chapter.

DOS AND DON'TS

There are certain fundamental rules that must be adhered to if any effort toward factory performance improvement is to produce significant and sustainable results. Some of these are summarized in the following list of dos and don'ts:

- Do seek to find the cause of and cure for factory performance problems. Don't seek simply to soothe the symptoms.
- Do focus your efforts on reducing variability and complexity while recognizing that there invariably will be resistance and skepticism within the typical organization toward virtually any change.
- Do seek the approval and, hopefully, involvement of senior management up to and including the firm's CEO; but don't expect this to happen without either a successful "marketing" effort (i.e., an effort taken to convince

management of the crucial importance of performance improvement) or solid and impressive evidence of some success (e.g., via a pilot study or relatively small-scale performance-improvement effort).

- Do select the right individual to lead the effort, that is, an individual with the education, experience, vision, and fortitude necessary to successfully implement the protocols necessary for performance improvement.
- Don't choose a performance-improvement leader whose only—or main—"strengths" are self-promotion, cronyism, and the propagation of unsubstantiated claims. In particular, don't select a leader who has a habit of taking credit for the ideas of others—particularly if they are the ideas of his or her subordinates. (While this would seem to be obvious, it is a sad fact that the culture inherent in some firms virtually ensures that such dodgy individuals are selected.)
- Do provide the necessary education and training for members of the group advancing the factory performance-improvement effort. Don't expect this to be accomplished in a day, week, or even a month. Do recognize that management may resist this crucial step and be prepared to advance arguments that support the need for education and training.
- Do recognize the obstacles that will be faced in any attempt to achieve significant and sustainable performance improvement within an organization that happens to be void of leadership and encumbered by a culture that not only resists but also fears change. While success may be possible in such an organization, it will require an immense amount of patience, passion, endurance, courage, and conviction.
- Don't separate improvement efforts in maintenance from those in operations. Do recognize that the same three enemies of manufacturing and the same three fundamental equations of manufacturing hold just as true for maintenance.[1]

---

[1] Factory performance-improvement efforts must focus on improving the *overall* performance of the entire factory. Dividing performance-improvement efforts into, say, those focused on cost reduction, those dealing with operational protocols, and those fixed on maintenance serve only to divide and degrade what could and should be a unified effort. Such a division is a recipe for failure.

- Do learn from the mistakes of others, including the fictional blunders described in the Muddle case studies.
- Don't limit your focus in terms of performance improvement to cost reduction. Do recognize that sustainable increases in profit, market share, and customer satisfaction should be a firm's primary goals.
- Do take the time and effort necessary to appreciate the history of manufacturing so as not to, as just one example, blindly accept the promises of the promoters of the latest manufacturing or management fad.

A CENTER FOR MANUFACTURING SCIENCE

The most promising and cost-effective path that may be taken to achieve overall performance improvement in a factory is to launch a center for manufacturing science. Obtaining approval to establish such a center may meet with resistance, though. It even may be perceived as a threat to the status quo. Ironically, although not surprisingly, the worse the performance of a firm's factories and the more lackluster its leadership, the harder it may be to gain management support for such an institute.

One way to gain the necessary backing for such a center is to mimic the marketing techniques employed by successful (at least in terms of receiving recognition and exorbitant fees) management consultants and motivational speakers. Even when they have little or nothing of substantive value to offer an organization—and even when there is no credible evidence whatsoever that their proposals will provide significant and sustainable improvement—they may, and often do, receive a rousing reception by top management, desperate for a quick and easy solution to the firm's woes.

By no means, however, am I suggesting that you lie, exaggerate, or make promises you know you cannot keep. What one needs to recognize, and what the most successful management consultants and motivational speakers provide an example of, is that you must package and market your product properly. The product in this case is the need for science in manufacturing and an appreciation of the powers of production-line protocols, that is, the benefits that may be accrued via exploitation of the third dimension of manufacturing.

Some managers may react negatively to any suggestion of a need for science, and thus rather than calling it a center for manufacturing science, you might want to call it a center for improved factory performance or possibly something even more alluring (and vague) such as a center for excellence in factory performance. *Excellence,* it would seem, is a noun that few are able to resist—just look at the titles of some of the most successful management books. Of course, as a last resort, you may have to name the unit the center for cost reduction. Hopefully, it won't come to this, but sometimes desperate measures are needed to deal with desperate situations.

You will be more likely to have such a center approved if you propose to begin relatively small, that is, no more than 6 to 12 personnel and modest requests for space and other support. The key to the ultimate success of such a center lies in two factors: (1) finding a corporate champion and (2) the qualifications and motivation of the team assembled. Before even considering seeking approval for such a center, both these matters should be considered.

The corporate champion should be someone who is approachable, open to new ideas, and hopefully, aware of the need for improvement in the firm's factories. The champion must have direct access to senior management and ideally the firm's CEO. In addition, the champion must recognize that the status quo will no longer suffice and that real, meaningful change is required. You must realize that if this champion is to stick his or her neck out for you, he or she must have confidence in your team's ability to obtain results. The champion wants to enhance his or her stature in the organization. Consequently, this person must be convinced—by your business plan and marketing skills—that the establishment of such a center will result in significant and visible results (e.g., evident all the way up to the CEO). In short, you want to convince this champion—perhaps without coming right out and saying it—that the center will serve to make the champion "look good."

Even before seeking out a corporate champion, you will be well served to establish a list of potential members of the center. Just as with the corporate champion, you must be able to persuade them that this endeavor will enhance their experience and job satisfaction and, if successful, advance their career prospects. This is something that you should be able to do if you are comfortable with the material presented in this book.

It should be noted that you or whomever the person may be who is advancing the idea of factory performance improvement and a center for excellence in manufacturing may not be the person best suited to lead such an effort. As Clint Eastwood said in his role as Dirty Harry, "a man's got to know his limitations."

If this person is a leader, then he or she may want to seek the position as leader of the group. If, however, the person is more comfortable with and more proficient at actually implementing the art and science of manufacturing, he or she might want to be the team's technical advisor or technical expert. In any event, if the individual is not comfortable with marketing the center, he or she should consider a position that will enhance the team rather than pursue what are possibly unrealistic aspirations.

Ultimately, the center's primary technical advisor (a.k.a. *internal consultant*) will determine the actual success of the center and the performance-improvement efforts. First, however, the need for the center must be marketed successfully, and later, its successes must be publicized (tastefully yet effectively). Hopefully, the leader selected for the center will be a person who gives credit where credit is due. If not, the morale, cohesion, and retention of the team assembled will be at extreme risk.

The remaining members of the team should be individuals selected on the basis of their motivation, willingness to work as a team, and technical qualifications. Perhaps the most important of these three attributes is motivation. Even if a team member is not an international expert in the politics, art, and science of manufacturing and its third dimension, he or she must have the motivation, discipline, and desire to become sufficiently well educated in these areas. He or she also should have an ancillary skill that directly supports the efforts of the center. Particularly helpful skills include

- Data collection, processing, and interpretation
- Simulation model development, exercise, and analysis
- Statistical analysis and the modeling of stochastic systems
- Operational research, systems engineering, and/or industrial engineering
- Knowledge acquisition and knowledge engineering
- Safety engineering and ergonomics
- Finance
- Human and organizational dynamics
- An appreciation of (and, as a distinct plus, actual factory floor experience in) the firm's factory operations, processes, and procedures

Ideally—and hopefully—the personnel comprising the team should be among the best and brightest of the firm's people. Unfortunately, there may be a temptation among management to assign personnel who at the moment "have no significant responsibilities" (these are, alas, code words to describe individuals in the firm who have shown so little motivation and skill as to not have made a positive impact on the organization—and are unlikely to do so in the future).

The final matter to be determined is that of the prioritization of the factory performance efforts to be undertaken. Management is typically, if not invariably, impatient for results, so the efforts initially undertaken should be those that (1) may be conducted relatively quickly, (2) are expected to produce significant results, and (3) are most likely to be concluded successfully. Efforts possessing such favorable attributes include

- Improvements in the factory starts protocol (e.g., starts declustering)
- Allocation of maintenance technicians to workstations
- Declustering of preventive maintenance (PM) events
- Removal or refinement of the production-line process steps
- Reduction of batch sizes

Ultimately, an educational outreach effort should be undertaken as early as possible in the life of the center. Short courses introducing the third dimension of manufacturing should be provided company-wide. Attendance should include employees at all levels and, in particular, plant managers, factory department managers, and senior factory engineers. Those attending these short courses should be encouraged to ask questions and explore possible causes of performance degradation within their factories and/or business processes. Rather than the presentation of a "course" in which the attendees are passive observers of PowerPoint slides, each person should be expected to participate.

In lieu of the establishment of a center for excellence in manufacturing, adoption and retention of the methods and philosophy of the topics contained in this book are less likely—at least until the present generation of managers moves on. However, there have been a few "grassroots" efforts that have achieved some degree of success.

In at least one factory within one multinational firm, the sheer doggedness and determination of a few individuals, coupled with the encouragement of a factory manager, resulted in a truly significant improvement in overall factory performance. While there has yet to be (at least at the time these words are being written) adequate appreciation of these efforts across the firm, there is at least some hope that over time the methods employed will receive the acceptance due them.

LEADERS VERSUS MANAGERS

One of the three enemies of performance is lackluster (or absence of) leadership. Of all the obstacles faced in factory performance improvement, this is by far the most difficult to overcome. Furthermore, while you might get away with informing corporate management that its factories are beset with unnecessary complexity and variability, it is unlikely that these same individuals will take kindly to any hint that they lack the attributes necessary for leadership.

First of all, however, it is important to recognize that leaders are not necessarily managers and managers are not necessarily leaders. Advancement to management positions within a firm are often a result of exceptional performance (or at least exceptional as perceived by one's superiors) in the management of a group, program, or project. For example, if a program is accomplished on time and within budget, this may be a sign of a person with good management skills, that is, the ability to assign personnel, schedule events, hold meetings, and overcome the obstacles common to almost any complex effort. While such a person may possess the necessary qualifications to be a manager, he or she may fall short of those required to be a leader.

The qualities that distinguish leaders from simply managers are not easy (and some would say impossible) to specify. Some leaders also may be competent managers. Many managers, however, lack the charisma, vision, and communication skills necessary to lead. As a consequence, some managers may focus their attention and efforts toward the development of mission plans, the invention of slogans, the establishment of goals (e.g., particularly short-term financial and market-share goals that appeal to Wall Street analysts), and the maintenance of their power and position (e.g., by means of currying favor with their superiors or populating the board of directors with those of a like mind).

Managers of this ilk are little more than departmental or corporate caretakers, advocates of the status quo, and they usually fail to inspire their subordinates and employees. While they can play an important role in the organization, they are more likely to push rather than to lead.

A leader, on the other hand, is willing to take reasonable risks and provide the means to accomplish the organization's goals (as opposed to simply voicing them and demanding results) and is able to inspire those he or she leads—as well as gain their loyalty. This does not necessarily mean that everyone in the organization will always agree with the vision and decisions of the leader. In fact, if this were the case, there probably would be no need for leadership.

It is also important to recognize that some leaders set their followers on precisely the wrong path (e.g., Hitler, Mussolini, Stalin, and Jim Jones of Jonestown infamy). They may have the charm, magnetism, and verbal gifts to attract a following, but they have chosen the wrong path. As such, it is vital that a leader have a vision that will enhance the organization (and possibly the state, the nation, and even the world) rather than advance a hallucination that leads to degradation and despair.

Another important, if not crucial, attribute of a successful leader is a willingness to listen to opposing viewpoints. Individuals in a position of power who refuse to hear anything with which they disagree guarantee a dysfunctional culture.

Saddam Hussein, for example, was able—through treachery, deceit, unbridled force, and outright terror—to become the president of Iraq. Like other tyrants before him, he had a vision—one in which he and a favored few prospered while most of his people suffered. While some have claimed that Hussein was a leader, consider his unwillingness to tolerate any views other than those he already espoused. If you were foolish enough to tell Hussein something that might well be true but counter to his personal belief, you risked having your tongue cut out or worse.

If you are to be successful in gaining the support of senior management and, hopefully, the CEO, you would be well advised to determine whether you are dealing with a manager or a leader. Managers are generally more risk adverse, less approachable, and more resistant to new ideas than leaders. If you recognize this and advance and package your proposal (e.g., for factory performance improvement or the initiation of a center for excellence in manufacturing) accordingly, your chances of obtaining the approval and resources necessary to initiate a meaningful improvement effort will be enhanced. Furthermore, if you are dealing with a manager rather than a leader, you must be able to (discreetly) convince him or her that the advancement of your proposal will benefit the manager directly.

On the other hand, if you are dealing with a leader, your plan for the advancement and acceptance of your proposal may have to be adapted to that which will appeal to this type of individual. While the typical manager will be more receptive to proposals that benefit his or her position and maintain or enlarge his or her sphere of influence, a leader may be more positively inclined to consider a proposal that will serve to achieve—or at least advance—the leader's vision.

This is not meant to imply that managers have huge egos and leaders are altruistic, selfless, and noble. Leaders can and often do have massive egos (e.g., George Patton, Napoleon, and Wellington). A difference between a manager and a leader is that a manager's ego is enhanced by maintaining and enhancing his or her power and position, whereas a leader's ego is most often fed by events that prove his or her vision to be the right one.

In the real world, you rarely, if ever, meet a pure manager or a pure leader. Most managers have at least some degree of leadership ability. Most managers—unlike some of those discussed in the Muddle Corporation case studies—are decent, honest, intelligent, and creative, but most exhibit the attributes and inclinations of a manager more so than those of a leader. Most leaders, on the other hand, have an ability to manage but prefer to lead.

These distinctions should be kept in mind if you have been given the opportunity to present your case for factory performance improvement to a firm's CEO and/or senior officers. They also should be considered when deciding on the leadership of the center for excellence in manufacturing.

EDUCATION AND TRAINING

As discussed in Chapter 1, it is my personal opinion that few academic or corporate training programs provide the education necessary to accomplish significant and sustainable improvement in factory performance. As just one example, consider today's academic programs in industrial and manufacturing engineering. While there are classes in a variety of useful topics, courses in classical industrial engineering and scientific management—the very foundation for the Toyota production system/lean manufacturing—became almost passé by the 1970s, replaced by classes that were considered to be more scholarly and more in step with modern times.

As a consequence, one could—and many did—graduate with a degree in industrial or manufacturing engineering with little or no appreciation of the contributions of Frederick Taylor, Frank and Lillian Gibreth, Harrington Emerson, Walter Shewhart, W. Edwards Deming, and Joseph Juran. Academic amnesia has reached such a point that it is believed by many that the Toyota production system emerged on its own and that the fundamental concepts of lean manufacturing are brand new.

Some have gone so far as to term this the *Topsy syndrome,* in reference to a quote from the unfortunate young slave girl in the novel, *Uncle Tom's Cabin*. When asked about her parents, Topsy replied, "I s'pect I just growed. Don't think nobody never made me." Her ignorance of her origin was sad but understandable. Ignorance of the history of manufacturing, however, should not be tolerated. The individuals and concepts that "made" the Toyota production system must be recognized, along with events, successes, and failures in the evolution of manufacturing.

As such, the history of manufacturing must be encompassed in any educational program for members of a factory performance-improvement team. No one finishing such a program should ever go away believing that the Toyota production system/lean manufacturing "just growed."

This book was written with the intention of providing material for support of the educational program advocated by me. Other excellent books exist (Hopp and Spearman, 2001; Levinson, 2002; Meyer, 1993; Standard and Davis, 1999) that may serve to augment this effort. The important point to be recognized, however, is that the educational program must provide attendees with a solid basis for an appreciation of the need for factory performance improvement and the means to achieve it in actual practice. Furthermore, it must be recognized that it takes more than cute slogans, clever diagrams, empty promises, and vague guidelines to achieve improvement.

Some firms have assumed, alas, that the educational process may be accomplished by a few hours or few days of exposure to PowerPoint slides. This naive notion (and mind-numbing approach) is a recipe for failure. While such a program—if limited to an hour or two and very carefully developed—might suffice to introduce senior management to the bare fundamentals of the art and science of manufacturing and its third dimension; it cannot possibly provide an adequate foundation for successful efforts in factory performance improvement.

One of the biggest mistakes management can make is to equate a course in manufacturing's third dimension to one in, say, how to position your computer's flat screen on your desk or how to fill out a purchase order. This sends a message to the firm's employees that management has equated the degree of knowledge required to improve factory performance to that necessary to perform an ordinary, mechanical, and mundane task. It also reveals the fact that any manager making such a statement has no experience in (and possibly no interest in) conducting any meaningful improvement efforts.

For those who will lead and/or directly participate in such efforts, a program extending over at least two weeks (and classes of six to eight hours per day) is recommended. Most important, the presentations/lectures must consist of something more than PowerPoint slide shows. Questions must be asked of attendees by instructors competent to do so—and instructors must be prepared to provide thoughtful and intelligent answers to the questions posed by attendees.[2] This requires something more—something far more—than the ability to hit the PowerPoint presentation advance key on the computer.

Ideally, quizzes and comprehensive exams should be part of the course. In an academic environment, this is expected. In a company training course, this practice may come as a rude shock—and even frighten away those who might benefit most from the material. To mitigate this reaction, it has been my practice to grade the quizzes and exams while assuring attendees that the grades will be kept absolutely confidential. This practice may be met with skepticism, but student feedback, in terms of graded papers, serves a vital, if not essential, role in education. It also discourages course attendees from "multitasking" (e.g., pretending to listen while surfing the Web on their laptops or responding to or sending e-mails) during course presentations. And trust me, without quizzes, exams, and class participation, multitasking will be the norm.

Once the formal course has been completed (or at least the majority of key points covered), attendees should be involved in meaningful and carefully planned exercises outside the classroom. Walks through the factory, development of process-step-centric flowcharts (possibly limited to just a segment of the production

---

[2] Individuals who "instruct" by means of nothing more than advancing the slides in a PowerPoint presentation are not instructors; they are readers.

line), examination and discussion of existing operating and maintenance specifications, collection of data such as the arrival times of jobs at a workstation, and assessment of such matters as factory start protocols are excellent first steps.

These relatively simple events often lead to observations and recommendations that serve to measurably improve factory performance—and lend credibility to the leverage potentially achievable by a simple change in factory protocols. Equally important, these experiences invariably serve to impress on attendees the importance and validity of the classroom experience.

Whatever the recommendations made by novice factory performance-improvement personnel, the instructor always should ask certain questions of the novice; that is,

- What motivated the recommendation (e.g., observation of data or observations on the factory floor)?
- Why does the novice believe the change will improve performance?
- Is the novice able to present a defense of his or her argument?
- What is the scientific foundation for the recommendation (e.g., how and why might it reduce variability)?

In other words, the novice must be capable of justifying any changes proposed. And this justification must be based on more than just hunches, past experience, and intuition. Until the novice is able to provide a solid, defensible rationale for any recommendations for changes, that person will remain a novice and should not be entrusted with the sole leadership of a performance-improvement effort.

WHAT ABOUT LEAN MANUFACTURING, ETC.?

If you have read and understood the material in the preceding chapters, you already should have the answer to the question, “What about lean manufacturing, Six Sigma, total productive maintenance, etc.?” In case you need a refresher, however, the answer will be summarized briefly in this section.

The fundamental concepts found in lean manufacturing, Six Sigma, total productive maintenance, theory of constraints, and a host of other proposals for factory performance improvement are encompassed within the material you have already (or should have already) covered. One difference, however, has been that of terminology. I have—definitely, intentionally, and unapologetically—avoided the use of Japanese words and phrases (e.g., *muda, mura, muri, kaizen, poke-a-yoke*, etc.) and instead have employed the words and phrases originated by the pioneers of scientific management (and only later translated into Japanese).

The primary difference, however, has been a matter of emphasis. Advocates of lean manufacturing, for example, typically focus their attention on a subset of the causes of complexity and variability and most often avoid any discussion of the culture and politics of the organization. Advocates of total productive maintenance focus their attention primarily on maintenance while mostly ignoring the fact that isolating maintenance protocols from nonmaintenance activities within a factory is counterproductive. This arbitrary separation may and often does build a wall between maintenance and operations that leads to decisions in one sector that have a negative impact on performance in the other.

The fact is that these concepts (i.e., lean manufacturing, Six Sigma, theory of constraints, total productive maintenance, and whatever might be the next "hot thing") represent a segmented reincarnation of the notions and practices embodied within their broader predecessors, that is, scientific management and classical industrial engineering. None of these "new" concepts is actually new, but each has a role to play, and each—if pursued properly—leads to improvements in factory performance.

One purpose of this book is to gather these fragments together under a unifying theme, that is, the politics, art, and science of production as practiced within the third dimension of manufacturing. Such unification should avoid the failure and disappointment rate now faced by these concepts when used in isolation. In short, the ideas and notions encompassed within lean manufacturing, theory of constraints, total productive maintenance, Six Sigma, and reengineering are important but are far more likely to achieve significant and sustainable improvement in factory performance if they are presented and employed in a unified fashion. Hopefully, this conclusion—and paragraph—will correct any misperception that I oppose the use of these methodologies. What I do oppose is their use in isolation and—in particular—where it is assumed that they are *the* answer.

OUTSIDE CONSULTANTS

As a final note on the elements required for a successful factory performance-improvement effort, we should consider the employment and role of outside consultants. An outside consultant, if chosen carefully, may serve to advance performance-improvement efforts significantly. Unfortunately, the wrong outside consultant—one who relies on slogans and has only a superficial comprehension of the complexity of a factory—can and will do more harm than good.

There is one thing that an outside consultant often can accomplish that is less likely to be achieved by the corporation's rank and file. This is the consultant's ability to find an audience with the firm's top management and perhaps even its CEO. There is a certain mystique about outside consultants. While they may be nothing more than a reasonably articulate person (with slides) who resides at least 100 miles away, they often are perceived to have knowledge that doesn't exist within the organization. This may be true, even though far more knowledgeable—and capable—individuals may reside in virtual obscurity within the firm's cubicles.

An outside consultant who is perceived as "the answer" is often able to convince management of the need to initiate certain efforts that simply would not be considered if proposed by a member of the firm's rank and file. Unfortunately, this also means that the consultant may be able to persuade management to undertake efforts that are counterproductive. Intriguingly, managers often seek the advice of outside consultants who have—never once—managed any effort or program of any meaningful size or complexity. This is akin to asking a stranger, a person who has never once played tennis, to teach your children how to become proficient at that game.

Consultants have the luxury of giving advice, receiving their compensation, and walking away. By the time their recommendations have been implemented (and have either succeeded or failed), they are on their way to their next consulting or speaking engagement.

This is not to imply that there are not some very capable consultants—individuals whose advice will advance the fortunes of their clients. The difficulty is, however, in determining whether a consultant will prove worth the expense or not.

One way to separate the good from the bad is to carefully evaluate the consultant's achievements, either real or perceived.

Selection of a consultant should require the same amount of time, effort, and investigation as selection of, say, a member of the board of directors or a senior executive. After all, a poor choice can inflict considerable damage to the firm. (Consider, for example, the immense damage that some "reengineering" consultants inflicted on their clients in the 1990s.) In short, some degree of skepticism is warranted in the selection and use of an outside consultant.

As someone who actually has served as a manager, I developed a list of red flags that I employed when considering the hire of an outside consultant. Anyone who exhibited the majority of these red flags was immediately shown the door. Those who exhibited several were assigned to the "suspect list." These red flags include

- Indications of the "one-trick-pony syndrome," that is, assertions that the particular concept being advanced is *the* answer—to whatever the situation
- Promises of quick-and-easy solutions
- Name dropping
- Record of hopping from one management or manufacturing fad to another
- Lack of knowledge of the history of manufacturing
- Lack of knowledge of or a reluctance to discuss alternate approaches
- Reliance on slogans and the proposal of vacuous guidelines
- Reluctance to participate in the actual implementation of any efforts that might be recommended
- Authorship of books and speeches long on promises but short on specifics
- Inability to cite specifics and get to the point when asked a direct question
- Unwillingness to admit to not knowing the answer to every question
- Unwillingness to spend any significant time on the factory floor and observing existing protocols
- Lack of awareness of or interest in the fundamental equations of manufacturing or an outright dismissal of their importance
- Inability to differentiate methods appropriate for synchronous factories from those designed for asynchronous facilities

- Excessive employment of such fuzzy and faddish terms as *excellence, robust, Pareto, strengths, actionable, bandwidth, multitasking, best practices, drill down, low-hanging fruit, learnings, paradigm shift, synergy,* and *teaming.*

CHAPTER SUMMARY

The successful accomplishment of any and all efforts devoted to improving factory performance requires more than the simple assignment of a team to a task (particularly if it is implied that the effort is to be in addition to the conduct of their existing activities). A corporate champion is a necessity in most firms, and the establishment of a center (e.g., center for excellence in manufacturing) is highly desirable. These basic elements, coupled with adequate education and training in the art and science of manufacturing, are the first steps toward success.

CASE STUDY 14: WHAT'S THE WEATHER LIKE IN FARGO?

Professor Aristotle Leonidas has invited Julia, Winston, and Dan to his home. He has as yet given no reason for the gathering. Each person seems lost in his or her own thoughts, and the only sound to be heard is that of a distant waterfall.

"Children," says the professor, breaking the silence, "what are your plans now that you're no longer employed by the Muddle Corporation?"

"Winston and I are considering starting our own consulting company," Julia answers. "We're convinced that the methods you've taught us could be put to use in supply-chain performance improvement. After all, as you told us, the underlying mathematical models of supply chains and factories are nearly identical. Reduce the complexity and variability in a supply chain and you've found the answer to improved performance."

"What about you, Dan?" asks the professor.

"I'm thinking about going to work for one of Muddle's competitors. Most of them operate just as inefficiently as Muddle. Maybe one of those firms will be more receptive to the methods you've taught us."

"Those are all good ideas," says the professor. "But I'd like to offer the three of you another option. One of Muddle's older, smaller factories has been sold to a group of investors. They're looking for people to get it up and running and to operate it efficiently. It doesn't produce Muddle's primary product, but it has been used to manufacture a supporting part. I think the new owners would look favorably on hiring any one or all three of you. Julia would make a fine plant manager, Winston would more than fill the bill as director of research and development, and Dan would be my choice for director of manufacturing. So, children, what about it?"

Julia is the first to answer. "Professor, it sounds like a great opportunity. But are you referring to Muddle's old Factory 1A, the one in Fargo?"

"I am," the professor answers. "Is that a problem?"

"I've never been to Fargo," Julia replies, "but I've seen the movie by the same name. As I recall, all it does there is snow. What about you, Winston, how do you feel about Fargo? Wherever we go, we go together."

"I happen to have been to Fargo," says Winston, "and they definitely have some hard winters. But they have some fine people. I'm all for applying for a job there if you are."

"Okay," says Julia, "let's get our résumés updated."

"And you, Dan, any interest in Fargo?" asks the professor.

"I'm game," Dan replies, "and I've already got my résumé updated."

"Excellent," says the professor, "but there's really no need to work on your résumés. As one of the investors in the Fargo facility, and having been authorized to hire its management, I can tell you that the three of you are, as of this moment, hired. Congratulations."

◆

Ben Arnold can barely contain his anger. Freddy Mertz, the new factory floor operations manager, is sweating profusely. The two junior plant managers have turned pale. The other meeting attendees have averted their eyes, sensing that they are about to witness a "train wreck."

"Freddy," says Ben, "let me stop you right there. You're telling us that the moves through the factory have been increased by almost 20 percent, our goal, but that you're convinced the figures are bogus?"

"That's right," Freddy replies, wiping his brow. "It seems that the people on the factory floor have figured out a way to increase moves while decreasing factory outs. They're using all kinds of tricks."

"What tricks?" screams Ben. "Who is doing this? I want them fired."

"First of all," says Freddy, "you'd have to fire pretty much everyone on the factory floor. Second, the tricks that are being used include everything from changing data to reworking perfectly good in-process units. They've also discovered that one clever way to increase moves is just to run those parts that take the least process time. Another thing that has happened is that we've now got a huge inventory of in-process units—too many to keep in the factory. The factory just can't handle the increased load, so people have been removing in-process units from the factory floor and storing them in the parts and supplies warehouses. Ben, it's out of control. This factory simply can't cope with your goal. It's just not physically possible."

"Don't go blaming me for your failure," shouts Ben. "You're fired."

Freddy Mertz simply shrugs his shoulders and leaves the conference room.

"You, Juan," says Ben, "as of this moment you're the new factory floor operations manager. I want you to . . ."

Before Ben can finish his sentence, Juan Gonzalez takes off his badge, tosses it to the floor, and follows Freddy Mertz from the room.

CHAPTER 15

Summary and Conclusions

THE IDEAL FACTORY

Based on the material covered in the preceding chapters, we may conclude that the ideal (i.e., utopian) factory should possess the following attributes:

- Single-piece, continuous process flow, as achieved by the elimination of
    - Batching or cascading
    - Priority jobs
    - Reentrancy
    - Need for rework
- Small, inexpensive, and relatively simple machines
- Adherence to the guideline (e.g., as set forth by Toyota) that the size of each machine should not be more than four times the size of the job it processes—unless otherwise dictated by the laws of physics
- Strict control and oversight over all links of the supply chain
- Impossible to produce a defect
- Impossible to induce an accident or injury
- Use of intelligent automation (e.g., the machines must be capable of monitoring and correcting their own performance)
- The employment of intelligent predictive maintenance

- A corporate culture that encourages human creativity and is receptive to change
- A single point of oversight with regard to operations and maintenance (e.g., an established center for factory performance improvement)
- C^4U-compliant operations and maintenance specifications
- The elimination of slogans
- The elimination of any temptation to chase fads and fashions

It also might be noted that if such a factory could be established, there would be no need for work-in-progress (WIP) management (i.e., job dispatch rules).

APPROACHING THE IDEAL

At this time, synchronous factories (e.g., automobile assembly lines and bottling plants) are closer to achieving the utopian goals just listed than are asynchronous production lines (e.g., semiconductor manufacturing facilities and certain pharmaceutical factories). Neither, of course, is likely to arrive at the ideal state in the foreseeable future—if ever. However, this should not deter us from establishing these goals and comparing an existing factory's performance with its ultimate but as yet unattainable form.

Some of the ideal performance measures may be reached—or at least significantly improved upon—by means of physical changes to the factory and its components. For example, progress in developing rapid thermal processing machines (e.g., for heat treatment) provides the means to reduce batch sizes, if not eliminate batching entirely. Higher-precision machines reduce the need for rework as well as inspection. And artificial intelligence (e.g., neural networks) could be and is being employed in the support of predictive maintenance. The importance of evolutionary and revolutionary changes in the physical components of a factory is undeniable. However, it still appears that the most promise toward achieving the ideal factory is through increased exploitation and understanding of the third dimension of manufacturing, that is, enhanced operating and maintenance protocols.

An awareness of the three enemies of factory performance, with at least some appreciation of the factors found in the three fundamental equations of manufacturing, serves as a guide toward improving performance by means of changes in protocols.

At all times we must measure and compare factory performance by means of metrics that are objective, normalized, and subject to oversight and audit. Such measures were introduced in Chapters 7 and 8.

The guidelines for enhancing factory protocols were laid out and illustrated in Chapters 10 through 12. The factors critical to the practical implementation (and acceptance) of the third dimension of manufacturing were covered in Chapter 14.

ZARA: A MANUFACTURING ROLE MODEL

I have presented examples of factories that have, to some degree, approached the ideal factory by means of exploitation of the third dimension of manufacturing. These included the Arsenal of Venice, the Ford Motor Company, and Toyota. One other firm should be considered because it provides a more current—and possibly even more intriguing—role model for any manufacturing firm seeking significant and sustainable improvement.

Spain's Inditex, a clothing manufacturer, has spent more than 30 years independently perfecting a strategy that incorporates virtually every fundamental concept proposed in this book. Zara is Inditex's wildly popular chain of clothing stores that serves as a retail link of Inditex's manufacturing efforts. The company, as of 2009, had either tied or surpassed The Gap as the world's largest clothing retailer. Since 2000, the firm had "nearly quadrupled sales, profits, and locations" (Capell, 2008).

Recall from Chapter 1 of this book that many of the manufacturing firms in the United States have either partially or wholly moved their factories to developing countries in a seemingly never-ending search for low wages and loose regulations. It also was mentioned that *approximately 96 percent of all clothing purchased in the United States is now produced outside the country*. Conventional wisdom among U.S. clothing manufacturers is that they have no choice but to produce apparel in developing countries. Inditex would seem to take issue with this assessment.

Inditex has turned conventional wisdom on its head. Rather than establishing factories in developing countries such as China, Inditex produces about half its clothing in factories in Spain and the nearby countries of Portugal, and Morocco. Factory workers in Spain, by the way, earn, on average, $1,650 per month in wages compared with $200 a month in China (and even less in other countries). Furthermore, Inditex supplies every market from its warehouses in Spain rather than locating those elements of it supply chain in other countries.

The approach employed by Inditex allows its Zara chain to receive new designs in its retail stores *within two weeks or less* while still remaining competitive. The best of Inditex's competition, on the other hand, require *eight months* between design conception and delivery to retail outlets.

As was the case with the Arsenal of Venice and Ford Motor Company, Inditex maintains control over every link in its supply chain—thus reducing complexity. This is coupled with the production of small batch sizes and the continuous pursuit of fast cycle times. In short, Inditex has explicitly or implicitly recognized and defeated two of the three enemies of manufacturing—complexity and variability. Furthermore, its owner (Amancio Ortega Gaona) has exhibited the attributes necessary to overcome the third and most difficult enemy—lackluster leadership. While Senor Ortega may be reclusive (there are only two known public photographs of him), the man definitely has a vision.

Ortega believes that market flexibility and minimal inventory levels are more important than cheap labor. While business schools—desperate for some means to describe the Inditex production system—cite the employment of lean manufacturing and just-in-time as the basis for Zara's success, they are missing the point. Zara's success is due almost entirely to the mitigation of complexity and variability coupled with the vision, persistence, and patience of Ortega. It's as simple as that.

CONCLUSION

If significant and sustainable factory performance improvement is to be achieved, it will require an approach that combines the art and science of manufacturing while considering at all times the impact of the culture and politics of the organization. While such concepts as lean manufacturing, Six Sigma, and total productive maintenance offer the potential to support such an effort, they should be part of a unified approach rather than implemented separately or in isolation. The emphasis of any manufacturing firm seeking significant and sustainable performance improvement should be on reducing complexity and variability—something that can only be accomplished by expertise in the science of manufacturing coupled with real leadership at the top.

CASE STUDY 15: FIVE YEARS LATER

The case studies and their characters that have been presented are, as has been noted, strictly fictional. They do, however, reflect to a degree the types of behaviors I and others have witnessed in some real-world factories. Fortunately, few real-world firms are as dysfunctional as the Muddle Corporation, and even fewer have managers the likes of which we encountered in the case studies. Most managers are, in fact, smart, honest, diligent, and dedicated. While some people do work their way up the corporate ladder through intrigue, back stabbing, and co-opting the ideas of others, they are—hopefully—in the minority.

Unfortunately, whatever the manufacturing firm and whomever might be in charge, it is all too common to encounter resistance to any proposal for changes that might require what may be considered "too much work." Some managers are at their most creative when it comes to making up excuses for avoiding change—of any type. Proposals for the inclusion of the third dimension of manufacturing in decision making and for changes in manufacturing protocols are often perceived to be "too difficult" mainly, if not solely, because they differ from the more typical "quick and easy" proposals traditionally delivered to management.

Sadly, it is far easier to convince management to accept a proposal that is based almost solely on sloganeering, vague guidelines, empty promises, and noble principles than one that requires serious thought and the conviction necessary for a change in corporate culture. It is also far easier for management unaware of the third dimension of manufacturing to purchase more machines and hire more personnel rather than to deal more cost-effectively with an underperforming factory. But, before you despair, let me assure you that it can be done.

There are some firms that, like Muddle, can never be convinced to change their ways. When faced with a problem, even a problem of immense proportions, some firms simply cannot seem to alter their ingrained response (e.g., note the plight of the "big three" American automobile manufacturers). Rather than admit that they have followed the wrong path, they continue to rely on the very same things that caused their problems in the first place; that is, they focus on cutting costs, closing plants, outsourcing, and—like Muddle—changing logos and slogans. These firms deserve the fate that awaits them, and there is little point in expending much time and effort in an attempt to convince them that there is a better way.

Fortunately, if you are patient (and knowledgeable about the third dimension of manufacturing), you should be able to convince management of a more receptive firm to give you at least an opportunity to prove your promises, be it via simulations or pilot studies. It does, however, require tact and diplomacy, attributes that are not emphasized nearly as much as they should be in the classroom.

Let's conclude the stories of the Muddle Corporation with a brief summary of the fortunes and misfortunes of some of the characters in the story five years from now.

* * *

Brad Simmons and Sally Swindel-Simmons

Brad and Sally's wildly popular book, *The Leadership Principles of the Donner Party,* reached number one on the list of best-selling business and management books. Sally is in great demand as a speaker and counts most of the Fortune 500 firms as her customers. Her fee, for an hour-long speech, is now $150,000. Brad, on his part, has discovered a passion for writing books on leadership. Buoyed by the success of *The Leadership Principles of the Donner Party,* he is hard at work on a new book, *The Leadership Principles of General George Armstrong Custer*. One may rest assured that it will receive a wide audience among managers desperate for advice that doesn't require any changes of consequence.

Julia Austen-Smith and Winston Smith

This happy pair was so effective in the startup and operation of the Fargo facility that they have now been placed in charge of three more factories. While they are now successful, their most prized possession is their job satisfaction.

Dan Ryan

Dan decided that before he could accept Professor Leonidas' offer, he should go back to school. He has now completed his dissertation and is ready to take on the duties as director of the center for manufacturing science at the professor's firm. One of the more interesting things he learned at university was the fact that he already knew considerably more about running a factory than his professors. He decided, however, to keep that fact to himself.

Professor Aristotle Leonidas

The good professor recently celebrated his ninety-third birthday. The professor claims, with a wink, that his longevity is due to a life free of complexity and variability. Prominent among the celebrants at his birthday party were Brad, Sally, Julia, Winston, and Dan.

Benedict "Ben" Arnold

Ben Arnold was promoted to the position of CFO of the Muddle Corporation. Considering the truly abysmal performance of Factory 7 under his reign, this actually may have been a good move for Muddle. Shortly after that promotion, Ben was hired by an alternative energy (wind turbines) firm to serve as its CEO. Between the stock options he received and his generous salary, Ben is doing quite well, thank you. One moral here is that the "bad guys" sometimes may win on the corporate battlefield.

Donna Garcia

Donna ultimately gave up looking for a comparable appointment at one of Muddle's competitors. At this time, her whereabouts are unknown.

Tommy Jenkins

Fewer than six months after his termination, Tommy managed to secure a factory manager position with one of Muddle's competitors. That factory's performance, since his arrival, has been on a downward spiral. Rumor has it that Tommy may be looking for a new job in the near future. He remains convinced, however, that there is no need for science in the operation of a factory.

Marvin Muddle

The Muddle Corporation continues to control the market for its primary product. This required, however, a fierce price war with its competitors and the use of some dubious marketing practices. The firm is faced with countless law suits in the United States and elsewhere citing unfair practices. Marvin Muddle remains the firm's CEO, although less and less is seen of him. There have been some unsubstantiated claims that Marvin, much like Howard Hughes in the 1950s, has become a recluse. At this point in time, it is rumored that he is working on the design of yet another new logo for his firm.

This page intentionally left blank

Bibliography

Alford, L. P., and H. R. Beatty, *Principles of Industrial Management*, revised edition, Ronald Press, New York, 1951.

Anderson, D., "Reducing the Cost of Preventive Maintenance," *www.plant-maintenance.com/articles/PMCostReduction.pdf*, no date.

Antil, P., "Baker Rifle," a description of the Brown Bess musket that may be found at *www.historyofwar.org/articles/weapon_brown_bess.html*, February 4, 2006.

Arthur, J., *Lean Six Sigma Demystified*, McGraw-Hill, New York, 2007.

Aurand, S. S., and P. J. Miller, "The Operating Curve: A Method to Measure and Benchmark Manufacturing Line Productivity," in *Proceedings of the Advanced Semiconductor Manufacturing Conference and Workshop,* September 10–12, 1997, pp. 391–397.

Baker, K. R., *Introduction to Sequencing and Scheduling*, Wiley, New York, 1974.

Barnes & Noble Classics, *Sun Tzu, The Art of War*, Barnes & Noble Books, New York, 2003.

Billings, R., and J. J. Hasenbein, "Heuristic Methods for Near-Optimal Wafer Fab Scheduling and Dispatching Using Multiclass Fluid Networks," in *Proceedings of the International Conference on Modeling and Analysis of Semiconductor Manufacturing,* Tempe Arizona, April 10–12, 2002, pp. 200–205.

Bodek, N., *Kaikaku: The Power and Magic of Lean*, PCS Press, Vancouver, WA, 2004.

Buzacott, J. A., and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems,* Prentice-Hall, Englewood Cliffs, NJ, 1993.

Capell, K., "Zara Thrives by Breaking All the Rules," *BusinessWeek,* October 20, 2008, p. 66.

Clason, T., *Fast Cycle Production,* 1st Books Library, Bloomington, IN, 2003.

Coman, K., *Industrial History of the United States,* Macmillan, New York, 1930.

Cooke-Taylor, R.W., *History of the Factory System,* Richard Bentley & Son, London, 1886.

Del Angel, C., and C. Pritchard, "Behavior Tests Six Sigma," *Industrial Engineering Magazine,* August 2008, p. 41.

Duffuaa, S. O., A. Raouf, and J. D. Campbell, *Planning and Control of Maintenance Systems,* Wiley, New York, 1999.

Encyclopedia Britannica, "Matthew Boulton," *Encyclopedia Britannica,* 2008.

Ford, Henry, in collaboration with S. Crowthers, *My Life and Work,* Garden City Publishing, New York, 1922.

Forrester, J. W., *Industrial Dynamics,* Pegasus, Waltham, MA, 1999.

Francis, R. L., and J. A. White, *Facility Layout and Location: An Analytical Approach,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

Galileo, G., *Discorsi e dimonstrazioni matematiche intorno a du nuove science,* 1638.

George, M. L., *Lean Six Sigma,* McGraw-Hill, New York, 2002.

Gilbreth, F. B., *Bricklaying System,* Myron C. Clark Publishing, New York, 1909.

Gilbreth, F. B., *Motion Study,* Van Nostrand, New York, 1911.

Goldberg, D. E., *Genetic Algorithms,* Addison-Wesley, Reading, MA, 1989.

Goldratt, E., and J. Cox, *The Goal,* North River Press, Great Barrington, MA, 1984.

Gross, D., and C. M. Harris, *Fundamentals of Queuing Theory,* Wiley, New York, 1998.

Hillier, F. S., and G. J. Lieberman, *Introduction to Operations Research,* McGraw-Hill, New York, 2005.

Hirano, H., and M. Furuya, *JIT Is Flow,* PCS Press, Vancouver, WA, 2006.

Hitomi, K., *Manufacturing Systems Engineering,* Taylor & Francis, London, 1996.

Hopp, W. J., and M. L. Spearman, *Factory Physics,* McGraw-Hill, New York, 2001.

Hounshell, D. A., *From the American System to Mass Production, 1800–1932,* John Hopkins University Press, Baltimore, 1984.

Ignizio, J. P., "Saturn S-II Telemetry Antennas: Ground Test Procedures," *North American Aviation S&ID Report,* Downey, CA, 1962.

Ignizio, J. P., "A Heuristic Solution to Generalized Covering Problems," Ph.D. dissertation, Virginia Tech, Blacksburg, VA, 1971.

Ignizio, J. P., "Solving Massive Scheduling Problems by Minimizing Conflicts," *Simulation,* March 1978, pp. 75–79.

Ignizio, J. P., "The Third Dimension in Manufacturing," technical paper, Pennsylvania State University, 1980.

Ignizio, J. P., *Introduction to Expert Systems,* McGraw-Hill, New York, 1991.

Ignizio, J. P., "The General Capacity Model," technical paper, Resource Management Institute, 1992a.

Ignizio, J. P., "Extensions of the General Capacity Model," technical paper, Resource Management Institute, 1992b.

Ignizio, J. P., "Metrics for the Evaluation of Factory Performance," technical paper, Resource Management Institute, 1997.

Ignizio, J. P., "Integrating, Cost, Effectiveness, and Stability," *Acquisition Review Quarterly* 5(1):51–60, 1998.

Ignizio, J. P., "An Optimization Model for the Declustering of Preventive Maintenance Scheduling," technical paper, Resource Management Institute, 1999.

Ignizio, J. P., "Fluid and Electromagnetic Network Simulation Models," technical paper, Resource Management Institute, 2000.

Ignizio, J. P., "Optimal Factory Tool Portfolios," *IEEE International Symposium on Semiconductor Manufacturing,* September 30–October 2, 2003a, pp. 31–34.

Ignizio, J. P., "The Implementation of CONWIP in Semiconductor Fabrication Facilities," *Future Fab International* 14, 2003b.

Ignizio, J. P., "Optimal Maintenance Headcount Allocation: An Application of Chebyshev Goal Programming," *International Journal of Production Research* 42:201–210, 2004.

Ignizio, J. P., "What Are the Alternatives to 450-nm Wafers?" *Future Fab International* 25:61–64, 2008a.

Ignizio, J. P., "Lean Manufacturing in the Semiconductor Industry: Proceed with Caution, Part I, *Fab Engineering & Operations* 3:37–44, 2008b.

Ignizio, J. P., "Lean Manufacturing in the Semiconductor Industry: Proceed with Caution, Part II, *Fab Engineering & Operations* 4:42–48, 2008c.

Ignizio, J. P., and J. N. D. Gupta, *Operations Research in Decision Making,* Crane Russak, New York, 1975.

Ignizio, J. P., and T. M. Cavalier, *Linear Programming,* Prentice-Hall, Englewood Cliffs, NJ, 1994.

Jensen, P. A., and J. F. Bard, *Operations Research: Models and Methods,* Wiley, New York, 2003.

Kennedy, J. B., and A. M. Neville, *Basic Statistical Methods for Engineers & Scientists,* IEP, New York, 1964.

Khade, S. B., and J. P. Ignizio, "A Heuristic for Batch Selection in Production Planning of Flexible Manufacturing Systems," in *Proceedings of Decision Sciences Institute,* Vancouver, Canada, 1990, pp. 363–366.

Levinson, W. A., *Henry Ford's Lean Vision,* Productivity Press, New York, 2002.

Liker, J. K., *The Toyota Way,* McGraw-Hill, New York, 2004.

Little, J. D. C., "A Proof for the Formula $L = \lambda W$," *Operations Research* 9:383–387, 1961.

Meyer, C., *Fast Cycle Time,* Free Press, New York, 1993.

Nolan, J., "Willow Run and the Arsenal of Democracy," *The Detroit News,* January 28, 1997 (also see *www.strategosinc.com/willow_run.htm*).

Paton, S. M., "Juran: A Lifetime of Quality," *Quality Digest,* August 2002.

Pegels, C., *Japan vs. the West: Implications for Management,* Kluwer-Nijhoff Publishing, Boston, 1984.

Popkin, J., and K. Kobe, "U.S. Manufacturing Innovation at Risk," *Council of Manufacturing Associations Report,* February 2006.

Robinson, C. J., and A. P. Ginder, *Implementing TPM,* Productivity Press, New York, 1995.

Roethlisberger, F. J., and W. J. Dickson, *Management and the Worker,* Harvard University Press, Cambridge, MA, 1939.

Sato, N., J. P. Ignizio, and I. Ham, "Group Technology and Material Requirements Planning: An Integrated Methodology for Production and Control," *CIRP Annals* 27(1):471–473, 1978.

Standard, C., and D. Davis, *Running Today's Factory,* Hanser Gardner Publications, Cincinnati, OH, 1999.

Stecke, K. E., and J. J. Solberg, "The Optimality of Unbalancing Both Workloads and Machine Group Sizes in Closed Queuing Networks of Multiserver Queues," *Operations Research* 33:882–910, 1985.

Taha, H., *Operations Research: An Introduction,* Prentice-Hall, New York, 2006.

Taylor, F. W., *The Principles of Scientific Management,* Harper, New York, 1911.

TECHNEWS, "Japanese Computer Chips Made at Too High Quality to Be Competitive on World Market," *www.technologynewsdaily.com/node/29238,* May 17, 2006.

Usher, A. P., *Industrial History of England,* Houghton Mifflin, Boston, 1920.

Van Zant, P., *Microchip Fabrication,* McGraw-Hill, New York, 2000.

Waddington, C. H., *OR in World War 2: Operational Research Against the U-Boat,* Elek Science, London, 1973.

Walton, M., *The Deming Management Method,* Berkley Publishing Group, New York, 1986.

War Production Board, Bureau of Training, "Training Within Industry Report," September 1945, p. 92.

Wikipedia, "Brown Bess Musket," *http://en.wikipedia.org/wiki/Brown_Bess.*

Wills, G., *Venice: Lion City,* Washington Square Press, New York, 2001.

Womack, J. P., and D. T. Jones, *Lean Thinking,* Free Press, New York, 2003.

Womack, J. P., D. T. Jones, and D. Roos, *The Machine That Changed the World: The Story of Lean Production,* HarperCollins, New York, 1991.

Wood, S. C., "The Impact of Single-Wafer Processing on Fab Cycle Time," in *Electronics Manufacturing Technology Symposium,* 1995, pp. 488–494.

Index

This page intentionally left blank

About the Author

James P. Ignizio received a Ph.D. in Industrial Engineering and Operations Research from Virginia Tech in 1971. He is a fellow of the Institute of Industrial Engineering, a fellow of the British Operational Research Society, and a fellow of the World Academy of Productivity Science.

Dr. Ignizio is the author of nine books and more than 350 publications, including over 150 peer reviewed papers in international professional journals. In 1980 he was awarded The First Hartford Prize by the U.S. National Safety Council for his contributions to this country's manned moon-landing program. Dr. Ignizio was also a fellow and senior research associate for the National Research Council from 1982–1983. In 2002 Dr. Ignizio was inducted into the Academy of Distinguished Alumni of the Department of Industrial and Systems Engineering at Virginia Tech.

Dr. Ignizio has held the positions of professor and chair: University of Virginia; professor and chair: University of Houston; professor: Pennsylvania State University; visiting professor: Naval Postgraduate School; visiting professor: U.S. Army ALMC; and distinguished adjunct professor: Helsinki School of Economics. His efforts in the fields of Industrial Engineering, Operational Research, and AI have received international recognition and his textbooks on these topics have been widely adopted in the United States and abroad. Dr. Ignizio's short courses on Operational Research, Industrial Engineering, Manufacturing Science, and Artificial Intelligence have been attended by several thousands of industrial, governmental, and military sector personnel over the past 30 years.

Prior to his 30-plus year academic career, Dr. Ignizio served as a program manager for the Apollo manned moon-landing mission. He was also deputy director of the Apollo/Saturn Integration Committee—with a focus on Fast Cycle Time. Dr. Ignizio has also served as a consultant to numerous industrial and governmental organizations in both the United States and abroad, including the U.S. Army, U.S. Air Force, U.S. Navy, NASA, SDI ("Star Wars"), Litton, Exxon, Texaco, Boeing, SAI, GRC, SRI, Bell Labs, Finland's Ministry of Economics, Quantas Airways, KAIST, Virginia Manufacturing Institute, Chase Bank, the Commercial Bank of Greece, and various semiconductor manufacturing firms.