5 Strategic Capacity Management

LEARNING OUTCOMES

After reading this chapter, you will be able to

LO 5.1 Explain what capacity management is and why it is strategically important.
LO 5.2 Exemplify how to plan capacity.
LO 5.3 Evaluate capacity alternatives using decision trees.
LO 5.4 Compare capacity planning in services to capacity planning in manufacturing.

LEARNING OUTCOME 5.1

CAPACITY MANAGEMENT IN OPERATIONS AND SUPPLY CHAIN MANAGEMENT

Check Your Progress

What is meant by capacity management in OSCM?

A dictionary definition of capacity is "the ability to hold, receive, store, or accommodate." In a general business sense, it is most frequently viewed as the amount of output that a system is capable of achieving over a specific period of time. In a service setting, this might be the number of customers that can be handled between noon and 1:00 P.M. In manufacturing, this might be the number of automobiles that can be produced in a single shift.

When looking at capacity, operations managers need to look at both resource inputs *and* product outputs. For planning purposes, real (or effective) capacity depends on what is to be produced. For example, a firm that makes multiple products inevitably can produce more of one kind than of another with a given level of resource inputs. Thus, while the managers of an automobile factory may state that their facility has 6,000 production hours available per year, they are also thinking that these hours can be used to make either 150,000 two-door models or 120,000 four-door models (or some mix of the two- and four-door models). This reflects their knowledge of what their current technology and labor force inputs can produce and the product mix that is to be demanded from these resources.

While many industries measure and report their capacity in terms of outputs, those whose product mix is very uncertain often express capacity in terms of inputs. For example, hospital capacity is expressed as the number of beds because the number of patients served and the types of services provided will depend on patient needs.

An operations and supply chain management view also emphasizes the time dimension of capacity. That is, capacity must also be stated relative to some period of time. This is evidenced in the common distinction drawn between long-range, intermediate-range, and short-range capacity planning.

Capacity planning is generally viewed in three time durat

Long range—greater than one year. Where productive resources (such as buildings, equipment, or facilities) take a long time to acquire or dispose of, long-range capacity planning requires top management participation and approval.

Intermediate range—monthly or quarterly plans for the next 6 to 18 months. Here, capacity may be varied by such alternatives as hiring, layoffs, new tools, minor equipment purchases, and subcontracting.

Short range—less than one month. This is tied into the daily or weekly scheduling process and involves making adjustments to eliminate the variance between planned and actual output. This includes alternatives such as overtime, personnel transfers, and alternative production routings.

The objective of **strategic capacity planning** is to provide an approach for determining the overall capacity level of capital-intensive resources—facilities, equipment, and overall labor force size—that best supports the company's long-term competitive strategy. The capacity level selected has a critical impact on the firm's response rate, its cost structure, its inventory policies, and its management and staff support requirements. If capacity is inadequate, a company may lose customers through slow service or by allowing competitors to enter the market. If capacity is excessive, a company may have to reduce prices to stimulate demand; underutilize its workforce; carry excess inventory; or seek additional, less profitable products to stay in business.

Capacity Utilization

The term **capacity** implies an attainable rate of output, for example, 480 cars per day, but says nothing about how long that rate can be sustained. Thus, we do not know if this 480 cars per day is a one-day peak or a six-month average. To avoid this problem, the concept of **best operating level** is used. This is the level of capacity for which process was designed and thus is the volume of output at which average unit cost is minimized. Determining this minimum is difficult because it involves a complex trade-off between the allocation of fixed overhead costs and the cost of overtime, equipment wear, defect rates, and other costs.

An important measure is the **capacity utilization rate**, which reveals how close a firm is to its best operating level:

$$\textit{Capacity utilization rate} = \frac{\textit{Capacity used}}{\textit{Best operating level}} \qquad \textbf{[5.1]}$$

So, for example, if our plant's *best operating level* was 500 cars per day and the plant was currently operating at 480 cars per day, the *capacity utilization rate* would be 96 percent.

$$\textit{Capacity utilization rate} = \frac{480}{500} = .96 \text{ or } 96\%$$

The capacity utilization rate is expressed as a percentage and requires that the numerator and denominator be measured in the same units and time periods (such as machine hours/day, barrels of oil/day, or dollars of output/day).

Economies and Diseconomies of Scale

The basic notion of **economies of scale** is that as a plant gets larger and volume increases, the average cost per unit of output drops. This is partially due to lower operating and capital cost, because a piece of equipment with twice the capacity of another piece typically does not cost twice as much to purchase or operate. Plants also gain efficiencies when they become large enough to fully utilize dedicated resources (people and equipment) for information technology, material handling, and administrative support.

At some point, the size of a plant becomes too large and diseconomies of scale become a problem. These diseconomies may surface in many different ways. For example, maintaining the demand required to keep the large facility busy may require significant discounting of the product. The U.S. automobile manufacturers continually face this problem. Another typical example involves using a few large-capacity pieces of equipment. Minimizing equipment downtime is

NOTES

Check Your Progress

1. Write in short about intermediate range capacity planning.
2. Write a short note about strategic capacity planning.
3. Write in short about capacity utilization rate.

essential in this type of operation. M
equipment to make M&Ms. A single packaging line moves 2.6 million M&Ms each hour. Even though direct labor to operate the equipment is very low, the labor required to maintain the equipment is high.

Capacity Focus

The concept of a **focused factory** holds that a production facility works best when it focuses on a fairly limited set of production objectives. This means, for example, that a firm should not expect to excel in every aspect of manufacturing performance: cost, quality, delivery speed and reliability, changes in demand, and flexibility to adapt to new products. Rather, it should select a limited set of tasks that contribute the most to corporate objectives. Typically the focused factory would produce a specific product or related group of products. A focused factory allows capacity to be focused on producing those specific items.

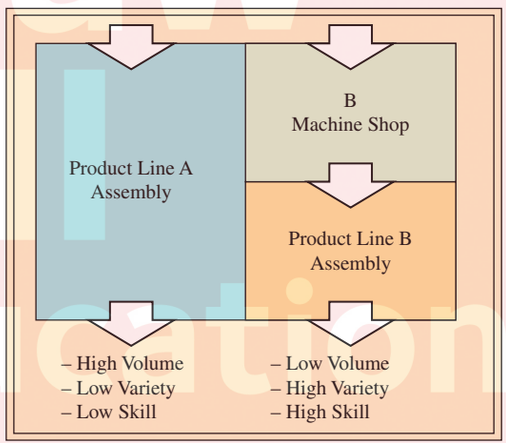
Check Your Progress

Give an example of plant within a plant.

The capacity focus concept can be operationalized through the mechanism of **plant within a plant**—or **PWP**. A focused factory (Exhibit 5.1) may have several PWPs, each of which may have separate suborganizations, equipment and process policies, workforce management policies, production control methods, and so forth, for different products—even if they are made under the same roof. This, in effect, permits finding the best operating level for each department of the organization and thereby carries the focus concept down to the operating level.

Exhibit 5.1 *Factories–Plant within a Plant*

This company needs to produce two different products. Product A is high volume and standard (there is no variation in how it is made). Product B is low volume and needs to be customized to each order. This plant is divided into three distinct areas that operate independently. The Product Line A area is a high-volume assembly line designed to produce A. B Machine Shop is an area where custom parts are made for product B. Assembly B is where product B is assembled based on each customer order. This factory, with its plants within a plant, can operate more efficiently than if both products were made with a single common production process.



Capacity Flexibility

Capacity flexibility means having the ability to rapidly increase or decrease production levels, or to shift production capacity quickly from one product or service to another. Such flexibility is achieved through flexible plants, processes, and workers, as well as through strategies that use the capacity of other organizations. Increasingly, companies are taking the idea of flexibility into account as they design their supply chains. Working with suppliers, they can build capacity into their whole systems.

Flexible Plants

Perhaps the ultimate in plant flexibility is the *zero-changeover-time* plant. Using movable equipment, knockdown walls, and easily accessible and reroutable utilities, such a plant can quickly adapt to change. An analogy to a familiar service business captures the flavor well: a plant with equipment that is easy to install and easy to tear down and move—like the Ringling Bros.–Barnum and Bailey Circus in the old tent-circus days.

Flexible Processes

Flexible processes are epitomized by flexible manufacturing systems on the one hand and simple, easily set up equipment on the other. Both of these technological approaches permit rapid low-cost switching from one product to another, enabling what are sometimes referred to as **economies of scope**. (By definition, economies of scope exist when multiple products can be combined and produced at one facility at a lower cost than they can be produced separately.)

Flexible Workers

Flexible workers have multiple skills and the ability to switch easily from one kind of task to another. They require broader training than specialized workers and need managers and staff support to facilitate quick changes in their work assignments.

Check Your Progress

1. Given an example of how an ice cream manufacturing plant can have flexible workers.
2. Write a short note on maintaining system balance.
3. Give an example of sharing capacity from the Indian manufacturing sector.

LEARNING OUTCOME 5.2

CAPACITY ANALYSIS

Considerations in Changing Capacity

Many issues must be considered when adding or decreasing capacity. Three important ones are maintaining system balance, frequency of capacity additions or reductions, and use of external capacity.

Maintaining System Balance

In a perfectly balanced plant with three production stages, the output of stage 1 provides the exact input requirement for stage 2. Stage 2's output provides the exact input requirement for stage 3, and so on. In practice, however, achieving such a "perfect" design is usually both impossible and undesirable. One reason is that the best operating levels for each stage generally differ. For instance, department 1 may operate most efficiently over a range of 90 to 110 units per month, whereas department 2, the next stage in the process, is most efficient at 75 to 85 units per month, and department 3 works best over a range of 150 to 200 units per month. Another reason is that variability in product demand and the processes themselves may lead to imbalance.

There are various ways of dealing with imbalance. One is to add capacity to stages that are bottlenecks. This can be done by temporary measures, such as scheduling overtime, leasing equipment, or purchasing additional capacity through subcontracting. A second way is through the use of buffer inventories in front of the bottleneck stage to ensure that it always has something to work on. A third approach involves duplicating or increasing the facilities of one department on which another is dependent. All these approaches are increasingly being applied to supply chain design. This supply planning also helps reduce imbalances for supplier partners and customers.

Frequency of Capacity Additions

There are two types of costs to consider when adding capacity: the cost of upgrading too frequently and that of upgrading too infrequently. Upgrading capacity too frequently is expensive. Direct costs include removing and replacing old equipment and training employees on the new equipment. In addition, the new equipment must be purchased, often for considerably more than the selling price of the old. Finally, there is the opportunity cost of idling the plant or service site during the changeover period.

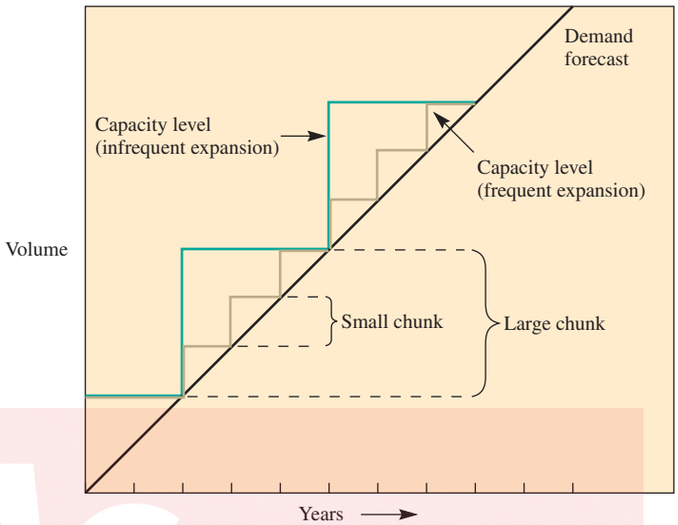
Conversely, upgrading capacity too infrequently is also expensive. Infrequent expansion means that capacity is purchased in larger chunks. Any excess capacity that is purchased must be carried as overhead until it is utilized. (Exhibit 5.2 illustrates frequent versus infrequent capacity expansion.)

External Sources of Operations and Supply Capacity

In some cases, it may be cheaper not to add capacity at all, but rather to use some existing external source of capacity. Two common strategies used by organizations are outsourcing and sharing capacity. An example of outsourcing is Dell Computer using a Chinese company to assemble its notebook computers. An example of sharing capacity is two domestic airlines flying different

NOTES

Exhibit 5.2 *Frequent versus Infre*



routes with different seasonal demands exchanging aircraft (suitably repainted) when one's routes are heavily used and the other's are not. A new twist is airlines sharing routes—using the same flight number even though the airline company may change through the route.

Decreasing Capacity

Although we normally think in terms of expansions, shedding capacity in response to decreased demand can create significant problems for a firm. Temporary strategies such as scheduling fewer hours or scheduling an extended shutdown period are often used. More permanent reductions in capacity would typically require the sale of equipment or possibly even the liquidation of entire facilities.

LEARNING OUTCOME 5.3

USING DECISION TREES TO EVALUATE CAPACITY ALTERNATIVES

Check Your Progress

Write a short note on decision trees.

A convenient way to lay out the steps of a capacity problem is through the use of decision trees. The tree format helps not only in understanding the problem but also in finding a solution. A *decision tree* is a schematic model of the sequence of steps in a problem and the conditions and consequences of each step. In recent years, a few commercial software packages have been developed to assist in the construction and analysis of decision trees. These packages make the process quick and easy.

Decision trees are composed of decision nodes with branches extending to and from them. Usually squares represent decision points and circles represent chance events. Branches from decision points show the choices available to the decision maker; branches from chance events show the probabilities for their occurrence.

In solving decision tree problems, we work from the end of the tree backward to the start of the tree. As we work back, we calculate the expected values at each step. In calculating the expected value, the time value of money is important if the planning horizon is long.

Once the calculations are made, we prune the tree by eliminating from each decision point all branches except the one with the highest payoff. This process continues to the first decision point, and the decision problem is thereby solved.

PLANNING SERVICE CAPACITY

NOTES

Capacity Planning in Services versus Manufacturing

Although capacity planning in services is subject to many of the same issues as manufacturing capacity planning, and facility sizing can be done in much the same way, there are several important differences. Service capacity is more time- and location-dependent, it is subject to more volatile demand fluctuations, and utilization directly impacts service quality.

Check Your Progress

1. How is capacity planning different in manufacturing and service sectors?
2. What do you mean by arrival rate in capacity planning?

Time

Unlike goods, services cannot be stored for later use. As such, in services, managers must consider time as one of their supplies. The capacity must be available to produce a service when it is needed. For example, a customer cannot be given a seat that went unoccupied on a previous airline flight if the current flight is full. Nor could the customer purchase a seat on a particular day's flight and take it home to be used at some later date.

Location

In face-to-face settings, the service capacity must be located near the customer. In manufacturing, production takes place, and then the goods are distributed to the customer. With services, however, the opposite is true. The capacity to deliver the service must first be distributed to the customer (either physically or through some communications medium, such as the telephone), then the service can be produced. A hotel room or rental car that is available in another city is not much use to the customer—it must be where the customer is when that customer needs it.

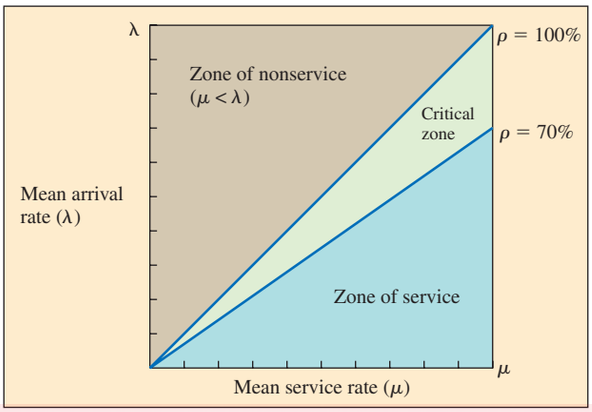
Volatility of Demand

The volatility of demand on a service delivery system is much higher than that on a manufacturing production system for three reasons. First, as just mentioned, services cannot be stored. This means that inventory cannot smooth the demand as in manufacturing. The second reason is that the customers interact directly with the production system—and these customers often have different needs, will have different levels of experience with the process, and may require a different number of transactions. This contributes to greater variability in the processing time required for each customer and hence greater variability in the minimum capacity needed. The third reason for the greater volatility in service demand is that it is directly affected by consumer behavior. Influences on customer behavior ranging from the weather to a major event can directly affect demand for different services. Go to any restaurant near your campus during spring break and it will probably be almost empty. This behavioral effect can be seen over even shorter time frames, such as the lunch-hour rush at a bank's drive-through window. Because of this volatility, service capacity is often planned in increments as small as 10 to 30 minutes, as opposed to the one-week increments more common in manufacturing.

Capacity Utilization and Service Quality

Planning capacity levels for services must consider the day-to-day relationship between service utilization and service quality. Exhibit 5.3 shows a service situation using waiting line terms (arrival rates and service rates). The term *arrival rate* refers to the average number of customers that come to a facility during a specific period of time. The *service rate* is the average number of customers that can be processed over the same period of time when the facility is operating at maximum capacity. The best operating point is near 70 percent of the maximum capacity. This is enough to keep servers busy but allows enough time to serve customers individually and keep enough capacity in reserve so as not to create too many managerial headaches. In the critical zone, customers are processed through the system, but service quality declines. Above the critical zone, where customers arrive at a rate faster than they can be served, the line builds up and it is likely that many customers may never be served.

Exhibit 5.3 *Relationship between*



Source: J. Haywood–Farmer and J. Nollet, *Services Plus: Effective Service Management* (Boucherville, Quebec, Canada: G. Morin Publisher Ltd., 1991), p. 59.

The optimal utilization rate is very context specific. Low rates are appropriate when both the degree of uncertainty and the stakes are high. For example, hospital emergency rooms and fire departments should aim for low utilization because of the high level of uncertainty and the life-or-death nature of their activities. Relatively predictable services such as commuter trains or service facilities without customer contact (for example, postal sorting operations) can plan to operate much nearer to 100 percent utilization. Interestingly, there is a third group for which high utilization is desirable. All sports teams like sellouts, not only because of the virtually 100 percent contribution margin of each customer, but because a full house creates an atmosphere that pleases customers, motivates the home team to perform better, and boosts future ticket sales. Stage performances and bars share this phenomenon.

Key Terms

Best operating level Output level where average unit cost is minimized

Capacity cushion Capacity in excess of expected demand

Capacity utilization rate Measure of how close the firm's current output rate is to its best operating level (percent)

Economies of scale Idea that as the plant gets larger and volume increases, the average cost per unit drops. At some point, the plant gets too large and cost per unit increases

Economies of scope When multiple products can be produced at lower cost in combination than they can be separately

Focused factory A facility designed around a limited set of production objectives. Typically the focus would relate to a specific product or product group

Strategic capacity planning Finding the overall capacity level of capital-intensive resources to best support the firm's long-term strategy

In Review

An operations and supply chain management view of capacity emphasizes the time dimension of capacity. Three time horizons are generally used: long range (greater than a year), intermediate range (next 6 to 18 months), and short range (less than a month). To distinguish between the absolute maximum capacity of the system (the highest output rate attainable) and the rate that is sustainable by

the system (that it can be run at efficiently and for a long period) is used. The utilization of the system is a measure of how close the system is operating relative to the best level.

When a producing resource, such as a manufacturing plant, gets larger and volume increases while at the same time average cost per unit of output drops, then the resource is exhibiting economies of scale. At some point the resource may be too large and the average cost will start to rise. This is when diseconomies of scale are a problem. Focused manufacturing plants are designed to produce multiple products using a concept called plant within a plant to improve economies of scale even though multiple products are produced in the same facility. This type of facility demonstrates the concept of economies of scope. Having capacity flexibility is often important to meeting the needs of a firm's customers.

From a strategic, long-term view, capacity additions or reductions come in chunks (fixed amounts). For example, an additional machine of a certain type is added to the existing pool of machines. Issues than involve how frequently and how much capacity is added or removed over time.

A useful technique for analyzing capacity problems is the decision tree. With this format the sequence of decisions are organized like branches in a tree. The potential consequences of the decisions are enumerated and evaluated based on their probability of occurrence and corresponding expected value.

Often services require that capacity be available immediately and that it be near where the customer resides. For example, a bank needs automated teller machines (ATMs) close to where customers want immediate cash, and enough of them so customers will not have to wait in long lines. Also, firms that offer services often need to deal with dramatic changes in customer demand over time (for example, the lunch-hour rush at a bank's drive-through window).

Multiple-Choice Questions

1. The ability to rapidly and inexpensively switch production from one product to another enables what are sometimes referred to as:
   a. Economies of scale
   b. Economies of size
   c. Economies of shape
   d. Economies of scope
   e. Economies of shipping
2. Capacity planning that involves hiring, layoffs, some new tooling, minor equipment purchases, and subcontracting is considered as which one of the following planning horizons?
   a. Intermediate range
   b. Long range
   c. Short range
   d. Current
   e. Upcoming
3. Capacity planning involving acquisition or disposal of fixed assets such as buildings, equipment or facilities is considered as which one of the following planning horizons?
   a. Intermediate range
   b. Long range
   c. Short range
   d. Current
   e. Upcoming
4. If the best operating level of a piece of equipment is at a rate of 400 units per hour and the actual output during an hour is 300 units, which of the following is the capacity utilization rate?
   a. 0.75
   b. 1.00
   c. 1.33
   d. 2.33
   e. 300
5. If the best operating level of a piece of equipment is at a rate of 400 units per hour and the actual output during an hour is 300 units, which of the following is the capacity cushion?
   a. 25 percent
   b. 100 units per hour
   c. 75 percent
   d. 125 percent
   e. 133 percent

6. The capacity focus concept ca
the following?
   a. Best operating level (BOL)
   b. Plant within a plant (PWP)
   c. Total quality management (TQM)
   d. Capacity utilization rate (CUR)
   e. Zero changeover time (ZXT)
7. The way to build in greater flexibility in your workers is to do which of the following?
   a. Pay higher wages to motivate a willingness to do a variety to tasks.
   b. Provide a broader range of training.
   c. Provide a wide variety of technology to augment workers skills.
   d. Institute a "pay for skills" program.
   e. Use part-time employees with specialized skills as needed.
8. Which of the following is not a step used in determining production capacity requirements?
   a. Forecasting to predict product sales
   b. Forecasting raw material usage
   c. Projecting availability of labor
   d. Calculating equipment and labor needs
   e. Projecting equipment availability
9. Which of the following models uses a schematic model of the sequence of steps in a problem and the conditions and consequences of each step?
   a. Probability indexing
   b. Johnson's sequencing rule
   c. Decision trees
   d. Activity system maps
   e. Decision mapping
10. Compared with a service operation, a manufacturing operation's capacity is which of the following?
   a. More dependent on time and location
   b. Subject to more volatile demand fluctuations
   c. Utilization more directly impacts quality
   d. Demand can be smoothed by inventory policies
   e. More capable of reacting to demand fluctuations

Key to MCQs

1. (d) 2. (a) 3. (b) 4. (a) 5. (b) 6. (b) 7. (b) 8. (b) 9. (c) 10. (d)

Discussion Questions

1. List some practical limits to economies of scale; that is, when should a plant stop growing?
2. Management may choose to build up capacity in anticipation of demand or in response to developing demand. Cite the advantages and disadvantages of both approaches.
3. Will the use of decision tree analysis guarantee the best decision for a firm? Why or why not? If not, why bother using it?
4. What are some major capacity considerations in a hospital? How do they differ from those of a factory?

Further Readings

1. Wright, T. P. "Factors Affecting the Cost of Airplanes." *Journal of Aeronautical Sciences,* February 1936, pp. 122–128.
2. Yu-Lee, R. T. *Essentials of Capacity Management.* New York: Wiley, 2002.