

Chapter 7: Capacity Planning and Management

Learning Outcomes

After reading this chapter you should be able to:

- Define and measure capacity and appreciate the factors that influence it.
- Assess the difficulties of matching capacity to demand.
- Evaluate and apply the different strategies for matching capacity with demand in the short, medium and long term.
- Analyse the impact of constraints and bottlenecks on a process and consider the Theory of Constraints.
- Outline the different strategies available for both manufacturing and service operations.
- Explain and apply some key forecasting techniques and understand the circumstances where each method is appropriate.

Opening Case Study: Waitrose and the Delia effect.

The supermarket industry has regularly used advertising and special offers to both increase over-all demand for their stores and also for individual and seasonal products.

Waitrose, the high end supermarket chain, recently abandoned its policy of not working with celebrities to endorse their products, by signing two world famous celebrity chefs on a three year deal. Delia Smith, a recipe writer, leading author and TV presenter joined forces with Heston Blumenthal, a maverick chef famed for egg and bacon ice-cream, cooking with dry ice and other such inventive creations in the kitchen, to front the advertising campaign.

The campaign involves Smith and Blumenthal filming a series of television adverts as well as appearing on billboards and in magazine features. The advertising strategy for Waitrose is to demonstrate a number of seasonal products being used in a recipe every week, in television commercials, backed up by a poster campaign and in house special offers. Most of the adverts are a short cookery program, with one or both of the cooks demonstrating the recipe, some adverts last the whole advertising break of three and a half minutes. Richard Hodgson, Waitrose's commercial director, said: "Every week will be a totally amazing new idea. Sometimes it will be a recipe, sometimes visiting a supplier or farmer, maybe even demonstrating a kitchen utensil," He promised that viewers would not be bombarded with some of Blumenthal's more wacky recipes. "Don't worry; we won't start selling snail porridge or liquid nitrogen. But I don't see why Heston couldn't demonstrate how a blow torch can be used to get the top of crème brûlée crispy."

Ocado, Waitrose's distribution partner has restructured their website to allow customers to order products direct from the online recipe. Waitrose have allocated prime locations on the end of the refrigerated isles with all the ingredients for the weeks campaign grouped together and on a special offer price.

It has not all been plain sailing however, the campaign came into criticism for the timing of a particular advert for Delia Smith's rhubarb and ginger brûlée. The advert was so early in the year, British rhubarb growers were unable to meet such large demand. The known phenomenon of the 'Delia effect', meant Waitrose sold enough of the plant for 61,000 desserts in four days alone, or the same quantity as it usually sells in 12 weeks. British growers said they simply could not meet the spike in demand. Much of Britain's rhubarb is grown in a small area of West Yorkshire between Bradford, Wakefield and

Leeds known as the 'Rhubarb Triangle'. Early in the year growers 'force' it by growing the plant in heated sheds, but there is limited capacity. Janet Oldroyd, of the Yorkshire Rhubarb Growers' Association, said this season's crop had been delayed by cold weather, further reducing output. A Waitrose spokesman said that sales of rhubarb were "so extreme" following the recipe being published online that it had to source from abroad.

A Delia recipe for fish risotto caused a backlash from even die hard Delia fans when it was described as 'vile' on the discussion section of the Waitrose website. One customer wrote, 'I bought the very expensive ingredients for the seafood risotto expecting to lay on a treat for my family. We all had one mouthful and gagged. It was disgusting to say the very least. The entire meal went in the bin and we all had toast instead.' Following the initial reactions to the recipe from news programs Waitrose appeared to be left with excess stock of the ingredients.

There is a requirement when planning these campaigns, for Waitrose to have a greater understanding of what may impact upon supply of the products and how much is likely to be demanded. This must be done with within the supply chain and the supermarket may need to be flexible within this plan to prevent further problems. Overall however Waitrose have seen an increase in demand for their products and are encouraging customers to be more seasonal with their selection of goods, this enables the suppliers to meet the supermarkets demand more easily.

Delia Smith and Heston Blumenthal to star in Waitrose ads [Harry Wallop](#), Consumer Affairs Editor
Published: 3:11PM GMT 03 Mar 2010 Telegraph online

Delia Smith causes rhubarb shortage By Stephen Adams
Published: 9:18AM BST 04 Apr 2010 Telegraph online

The risotto rebellion: Anger of Delia fans left gagging on her 'vile' seafood recipe Jo McFarlane 3rd
May 2010 Daily Mail online

1. Introduction

The case demonstrates how complicated it can be to deal with the unpredictability of demand and how difficult this can be to match this to the supply of goods and services and still have an efficiently run operation. The difficulty of meeting customer shopping patterns (**demand**) with what the suppliers and therefore supermarket can provide (**capacity**) is a complicated timing issue (**forecasting**). Providing the right amount of good at the correct time requires careful planning and investment decisions.

There are 3 things an operation must consider when looking at production of goods and services:

Capacity:

- How much can they make?

Demand:

- How much does the market require?

Forecasting:

- How are they going to match what is required with what they can make without wasting resources?

These factors all impact upon each other; in order to fulfil the requirements of a market, an operation must be able to evaluate what capacity it can provide and the options it has to increase or decrease this.

It is not just a case of offering the amount of goods or services wanted; it is also an issue of timing and when these are required. This can give an operation a number of problems when attempting to plan capacity.

An operation may experience many changes and fluctuations in the demand for its products or services and must be able to maximise opportunities of demand without having excess capacity. For example, the daily demand for a restaurant often peaks in the evening, but this may not be the only pattern in demand, the weekends may be busier than a weekday and there may be other events like Valentine's Day or the Christmas period where demand may be expected to increase.

The provision of capacity is a decision that should not be taken lightly; for once implemented it cannot be easily reversed and the operation is then left to manage the choice it has made.

This chapter considers how capacity can be provided, adjusted and managed in order to satisfy the demand of the consumer and meet the objectives of the operation as efficiently as possible.

2 Capacity Management

Capacity management affects all areas of an operation. Capacity **measures** the rate that the operation can transform inputs into outputs. Capacity is about the **quantity** of a product or service that can be made within a **given time period**. This, for example, could be;

- The number of passengers per flight on an aeroplane
- The number of patients that can be seen in a surgery session at a doctors
- The number of mobile phones that can be produced a week by a factory.

It is defined as the number of units (goods and/or services) an operation can produce over a given time period, under normal working conditions, where no additional resources are deployed.

Capacity is usually measured in convenient units such as litres per hour or passengers per taxi. For instance, a domestic tap may be able to deliver 20 litres per minute of water; a bus may have a capacity of 53 passengers, a football stadium may be able to seat 50,000 spectators or a McDonalds may be able to serve 600 customers per hour.

In many instances capacity may be simple to calculate, however more difficult questions might be:

- How many fire engines should an airport have on standby?
- How many operations should a surgeon schedule?
- What service level should be offered to broadband customers?

When planning capacity there are always two sides to consider: - firstly there is the demand – the amount of the product or service that might be wanted; and secondly there is the provision of the good or service. In providing products or services the operation must evaluate the costs involved and the trade off between satisfying customers and the costs of production. Having too little capacity to respond to customer demand may mean missed opportunities and annoyed customers, however under-utilised capacity is a waste of resources resulting in higher costs.

2.1 Adjusting Capacity

Discrepancies between the capacity of an organisation and the demands of its customers result in inefficiency, either in underutilised resources or dissatisfied customers. The former may be a serious cost but the latter may result in lost sales, lost customers and potentially loss of reputation. The ability of the operation to adjust the key resources will minimise these discrepancies.

Some capacity changes can happen almost instantly, others may take longer time to put in place. The capacity of an operation is a complicated mix of resources. These resources are inputs to the process that allow capacity to be expanded or contracted, by changing the inputs into the process. How flexible the resource is depends on how quickly it can be altered.

Capacity can be increased using a number of methods which involve adjusting the resources and inputs into an organisation such as:

- Introducing new approaches and materials
- Increasing the number of service providers or machines
- Increasing the number of operational hours

- Acquiring additional facilities.

Decreasing capacity can be more difficult or expensive; it tends to rely on the operations ability to sell or reduce resources as cost effectively as possible. There are usually costs involved in reducing resources, for example if the resource is staff there may be redundancy costs, or closing facilities may incur significant costs.

Therefore the decision to alter capacity has to be taken carefully in line with future predictions of demand.

2.2 Capacity Constraints

A constraint on capacity is a resource that is less capable, of increasing its throughput over the given time period, than other parts of the operation. A number of machines may be in sequence on a manufacturing line yet one may not be able to process as many units per hour as the other machines. The capacity will be constrained by this under producing machine and this may create a 'bottle neck' in the process. By increasing the capacity of this machine the capacity of the overall facility will also increase.

Capacity is always constrained by the lowest producing part of the process. In layman's terms an operation will ***'always go at the pace of the slowest walker'***. Identifying a restrictive part of the process and adding resources that can increase its capacity will improve the overall capacity of the operation.

The **resource mix** that can be potentially constraining to an operation could include;

- **Staff/Skill levels:** Staff can be trained over time to be more flexible in their contribution to the process. The operation can benefit from the learning curve, where a new employee can become more efficient at a given process and therefore be quicker at their job, which can increase the capacity of the operation.
- **IT facilities/Technology:** This can be a small or very significant improvement to a process. The investment in ICT can reduce process time or even completely change the nature of the process itself. Online banking has been a significant improvement in the finance sector by reducing the number of staff required to process a transaction and therefore massively increasing the capacity of the bank to deal with its customers.
- **Materials availability:** A change in the supply of raw materials can increase the capacity potential of an operation. If there is a restriction in availability of materials or a timing problem and this is released, the capacity could be improved.
- **Product or service mix:** Adjustments in the other products or services made by the facility can restrict the capacity of the operation. This is because different products and services may use different quantities of resources per unit; therefore a change in the product mix may result in a change in capacity.
- **Storage:** This can affect the capacity of an operation if there is a resource constraint that is affected by timing in the process. If the operation has the ability to store work in progress or finished goods it can improve the capacity of the process in the short term. The swings and fluctuations in demand can be mitigated by the ability to store products and allow the full capacity of the operation to flow.
- **Working schedules and access to facilities:** This can also dictate the full availability of capacity. A lecture theatre that can accommodate 100 students at a time could operate beyond a standard working day; however both staff and students may have an issue regarding 6am lectures!

These factors should be considered in a short-term, medium-term and long-term time frame to establish their ability to be changed over time. A short term strategy for expanding capacity in a cafe,

would be to put a few extra tables outside or extend the staff working hours to cope with the extra demand, in the medium term the cafe owner would have more options available to increase the capacity, such as hiring more staff or having additional cooking facilities in the kitchen to cope with extra demand. In the long-term the possibilities can be much greater, the premises could be expanded, better equipment, more staff and so on. The options available to an operation are greater the more time it has to plan them.

2.3 Theory of Constraint (TOC)

The theory of constraints was first proposed in 1986 by Goldratt. The theory is the practical results of Goldratt's work on 'how to think'. TOC is a philosophy that suggests that any system must have **at least one constraint** otherwise it would generate an infinite amount of output and that constraints generally determine the pace of an organisation's ability to achieve its goal which is profit.

Goldratt emphasises that constraints pose a significant threat to the wellbeing of an organisation and must be identified. He suggests that constraints may be labour availability, staff skills, machine availability, and capital or time available. They may however be more difficult to identify such as; organisational policies, guiding principles or rate of innovation.

He identifies that there is rarely an equal flow of work within each work centre in a process. The constraint should therefore be the control of the pace of the process. This theory reduces the emphasis on maximising all resources within the process and prioritises the management of the bottleneck. The theory he advocates is called 'drum, buffer, rope' where the bottleneck is the 'drum' which marks the time through the process – due to insufficient capacity this should be working the most. The 'buffer' principle is required to make sure that the bottleneck is never short of work and therefore the front end of the process should stockpile inventory to maximise output. The 'rope' element is the communication device to make sure the front part of the process does not overproduce.

Goldratt's five focusing steps

The theory identified a process to follow in order to free the system from the bottleneck that slowed it down. By following the steps the operation identifies and clears the blockage, this will then in turn reveal a new bottleneck and the 5 steps can start again.

- 1 identify the systems constraints
- 2 exploit the systems constraints
- 3 subordinate everything else to the above decision
- 4 elevate the systems constraints (identify the next constraint)
- 5 if in a previous step, a constraint has been removed, go back to step 1 but do not allow inertia to become the systems constraints.

Goldratt advises that any constraint having been identified is only transitional. As this constraint is exploited, another will appear in its place. Without identifying the real constraints, Goldratt suggests that management may not be able to find the real causes that restrict capacity so will take actions to work around the problem rather than solve the real cause.

Constraint analysis is a subject that is larger than the subject of Capacity Management. However it does offer an important perspective on the question; **'is all capacity equally important?'**

Short case: Wind Energy; where's the constraint?

According to the China Wind Energy Association, the power output generated by the farms was not meeting expectations. Over 20% of the power created was not reaching the national grid. The reason for this under performance of the system was that new wind farms and turbines set up and working but were not attached to the grid for around the first three to four months. The utilisation rate of mainland wind farms in China, much lower than comparable projects operating in the UK or USA

While differences in wind resources could partly explain the disparity, insufficient investment in grid connections played a key part. The problem appears to be that the regions where wind resources are most abundant tend to have the weakest grid. In addition, some wind farms are generating more power than initial estimates and their output has been rejected by the local grid companies because of concerns it would overload their systems.

Wind farms are generally of smaller generation capacity than coal-fired power plants and are less stable in output since they are subject to variable wind speed. As a result, the return on investment from wind power is lower for grid operators.

(Based on 'Wind power loses puff - infrastructure to tap captured energy is lagging headlong capacity expansion' - South China Morning Post, July 13, 2009

Questions

1. What resources are needed to reduce or remove the current bottleneck?
2. Where could the next bottleneck occur if this bottleneck is reduced or removed?

The key point is:

A constraint will restrict the capacity of the operation, by reducing or even removing this constraint capacity will increase. A constraint can be reduced by increasing the resource in the bottleneck.

3 Measuring capacity

When measuring capacity the unit of measure can be either an input or an output to the process. The key is to take the most logical unit that reflects the ability of the operation to create its product or service. However, where the input is more complicated to measure, such as machine hours on a process layout, then output is a more suitable measure. The unit of time could be a minute, an hour, a day or a week, or whatever time scale fits the operation, but the unit of output and time scale needs to **be consistent**.

3.1.1 Input measures of capacity

When using input measures of capacity, the measure selected is defined by the key input into the process. Where the provision of capacity is fixed, it is often easier to measure capacity by inputs, for example; rooms available in a hotel or seats at a conference venue. Input measures are most appropriate for small processes or where capacity is relatively fixed, or for highly customised or variable outputs such as complicated services.

3.1.2 Output measures of capacity

The output measures count the finished units from the process such as mobile phones produced in a day or cars manufactured per week. This measure is best used where there is low variety in the product mix or limited customisation.

Table 7.1 An example of possible input or output measures.

Process	Input capacity measure	Output capacity measure
Music festival	Square metres of land	Number of festival attendees
Hotel	Rooms available	Number of guests per week
Car Manufacturing plant	Machine capacity	Cars produced per month
Milk Bottling Plant	Machine hours available	Bottles filled per day
Lecture theatre	Number of seats available	Students on courses
Wedding planning service	Consultants available	Weddings per season

Capacity can be measured from looking at the operation as a whole and then calculated on the resources and facilities available and process time. Table 7.1 shows the alternatives that can be used for input/output measures.

For example, the measure of output capacity could be cars per shift or tonnes per hour or customers per day. However, the capacity of a surgeon or a University Professor may not be measured in this manner. In these cases, capacity could be shown in the form of working hours per week.

A simple formula for capacity can be:

$$\text{Capacity} = \frac{\text{Time available}}{\text{Time of task}}$$

For example, a service provider works an eight hour day, takes two fifteen minute coffee breaks and has a half hour lunch break. The time available for work is seven hours per worker per day.

If this particular worker was a fitness instructor and he spends 70 minutes with each customer (10 minutes for the consultation and booking and 1 hour for the gym session), how many clients could the instructor process during a five day week?

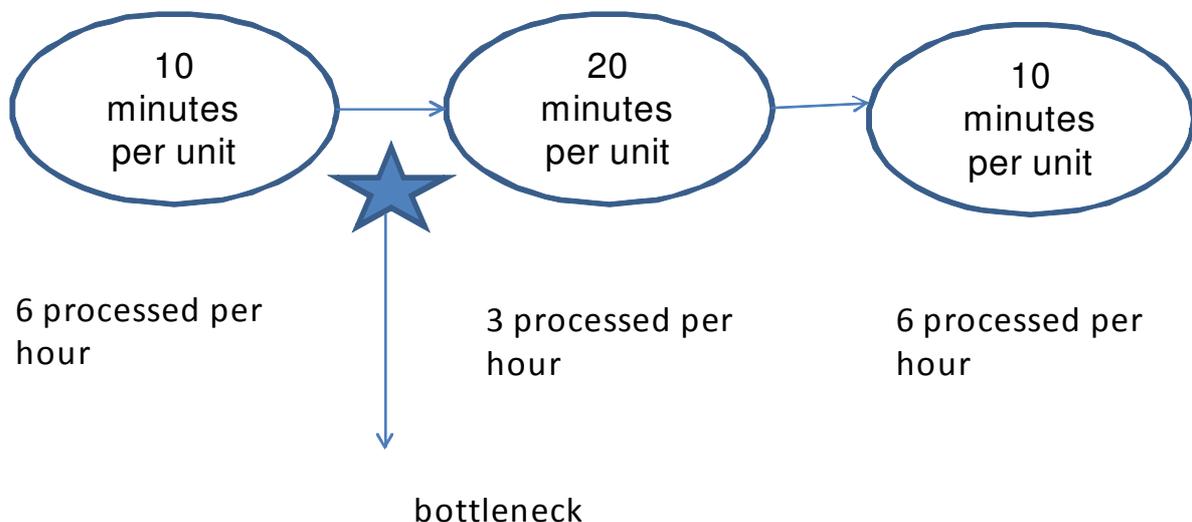
$$\frac{(7 \text{ hours per day} \times 60 \text{ minutes per hour}) \times 5}{70 \text{ minutes per client}} = 30 \text{ clients per week}$$

30 Clients per week can be expressed as the capacity of the fitness operation.

This is a simplified measure as it presumes that the fitness instructor doesn't have time off sick or do any other activities such as maintain the gym equipment or diversify into other areas such as taking classes. Most processes will not have just one activity; many will have interlinking processes with different capacity constraints on each.

Here the operation will have to consider the capacity of the whole process and not individual constituent processes. Also the individual process durations may differ. If the first part of the process takes 10 minutes but stage 2 takes 20 minutes and stage three takes 10 minutes then a backlog will appear at stage 2.

Figure 7.1 Shows the bottleneck point in a simple process



The diagram shows that the output of a process will be constrained by the slowest point. This is referred to as a 'bottleneck' in the process.

However it is not always possible to accurately predict how long each stage is actually going to take. A hair dresser, for example, may allocate thirty minutes to each haircut, forty minutes to each hair colorant and ten minutes to styling, but individual customers may take more time and others less. In such circumstances it may not be possible to accurately locate the bottleneck in variable processes.

This shows an important feature of capacity planning, assumptions must be made as to what the process is capable of in order to understand the output of the operation. However, although assumptions are needed to plan the process, often in reality these assumptions can be found to be inaccurate.

3.2 Defining Capacity

Capacity is defined under 3 categories; design capacity, effective capacity and actual capacity. The operations utilisation of resources and the efficiency of its processes can then be calculated using these.

3.2.1 Design Capacity

This is a theoretical number and not one that is applied to the daily production of an operation. Design capacity is the output that an operation can produce continuously, at **maximum rate without stopping** for any shift changeovers, maintenance or any other delays. What the process is capable of producing under perfect conditions. In some cases this might be interpreted as maximum capacity.

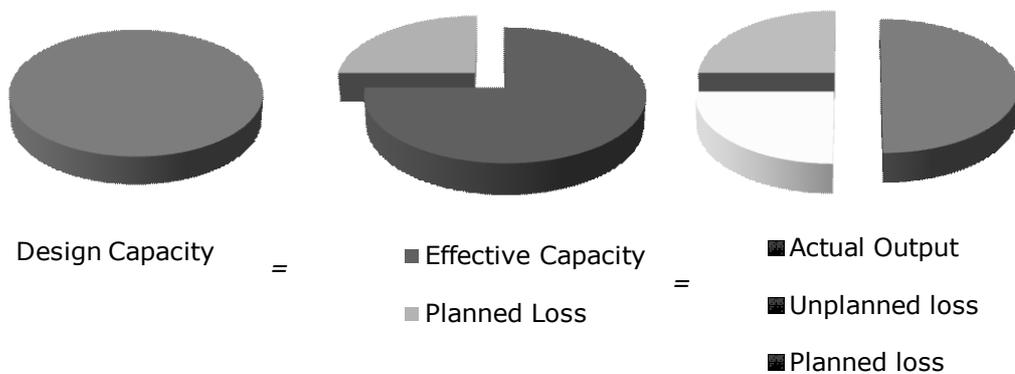
3.2.2 Effective Capacity

This considers how the operation will run on a long term basis, how it will be staffed and how it will be maintained. All **planned** stoppages under the normal working time frame are taken into consideration. This can also be known as *available capacity*. These stoppages may include shift changeovers, lunch breaks, set up times and many other operational factors.

3.2.3 Actual capacity

This is the same as effective capacity but contains **unplanned losses** as well as planned ones. These could include poor work rate, absenteeism or new staff training for example.

Figure 7.2 The proportions of capacity measures



Actual output plus unplanned losses is the same as effective capacity. Therefore the operation which is working its assets efficiently is minimising unplanned losses.

3.3 Efficiency and Utilisation Calculations

For the efficient use of the resources available, efficiency is output shown as a percentage of available capacity.

$$\text{Efficiency} = \frac{\text{actual output}}{\text{effective capacity}}$$

For an operation that has been well designed, there will be minimal planned losses. This allows the resources to be used to the best of their ability. Capacity utilisation is the measure of how much of the available capacity is used. Utilisation is output shown as a percentage of the facilities or designed capacity.

$$\text{Utilisation} = \frac{\text{actual output}}{\text{design capacity}}$$

For example if the fitness trainer in the previous example only had 24 clients who arrive at their appointments on time, the calculation would be:

$$\text{Utilisation} = \frac{24}{30} \times 100 = 80\%$$

Therefore the utilisation rate is 80%

These measures of capacity can tell an operation how well they are utilising their resources and how efficient the manufacturing process is.

Worked example 1: Calculating capacity measurements

You are managing a group of 10 Electricians. These individuals undertake in-home servicing of electrical systems and are called by telephone for either emergency or pre arranged visits. They charge a minimum call out fee that covers the first 15 minutes of their visit plus travelling time. Beyond the first fifteen minutes they charge in minimum blocks of 15 minutes plus any materials that might be necessary to carry out the job. The average call out takes 1 hour.

The workers usually are available for eight hours a day but with 2 coffee breaks of 15 minutes each and a half hour lunch break, they actually work a 7 hour day. Taking time off and illness into account reduces the electricians' available time by 20%. This means the 7 hours per day is reduced to a 5 hour and 36 minute day (5.6 hours)

If actual work is only 200 hours billed in the week then (a) What is the capacity utilisation of the team? (b) What is their efficiency?

Approach: First, you need to calculate (a) the design capacity and (b) the effective capacity. Then use the actual output given above to calculate the capacity utilisation and efficiency.

Solution:

$$\text{Design Capacity} = \frac{10 \text{ workers} \times 7 \text{ hours per day}}{1 \text{ hour per customer call out}} \times 5 \text{ days of a working week}$$

$$\text{Design Capacity} = 350 \text{ customers per week}$$

$$\text{Effective Capacity} = \frac{10 \text{ worker} \times 5.6 \text{ hour day}}{1 \text{ hour per customer call out}} \times 5 \text{ days of a Working week}$$

$$\text{Effective Capacity} = 280 \text{ customers per week}$$

;

$$\text{Efficiency} = \frac{\text{actual output } 200}{\text{effective capacity } 280}$$

Then the efficiency of the process is 71%

$$\text{Utilisation} = \frac{\text{actual output } 200}{\text{design capacity } 350}$$

The Utilisation of the process is 57%

Note: for this example, **capacity** is determined by the difference between actual hours of work measured against available or attended hours multiplied by an efficiency factor that takes into account current conditions.

3.4 Takt or Cycle Time

In the above example, we suggested that the time to service each customer was 1 hour. This is often referred to as the Takt time.

Takt or cycle time can be defined as the maximum time per unit allowed making a product or providing a service in order to meet the demand. It is derived from the German word Taktzeit which translates to **cycle time**.

Takt time usually determines the output rate for manufacturing lines. For instance in automobile manufacturing cars are assembled on an assembly line, and are moved from station to station after a certain time – this is the takt time. In a fast food restaurant the service time for each order would also be called the takt time.

Takt (or cycle) Time can be first determined with the formula:

$$T = \frac{T_a}{T_d}$$

Where:

T = Takt time, e.g. [minutes of work / unit produced]

T_a = Time available to work, e.g. [minutes of work / day]

T_d = Time demand (customer demand), e.g. [units required / day]

Available time is the amount of time available for work to be done. This excludes break times and any expected stoppage time (for example scheduled maintenance, team briefings, etc.)

Worked Example 2 Calculating Takt time

An assembly line has a total of 8 hours available time in a shift. The employees working the shifts have a half hour lunch break and two fifteen minute tea breaks each shift. The machines also require ten minutes per shift in basic maintenance. At the start of each working day, the supervisor spends ten minutes talking to the staff and setting goals for the shift ahead.

Calculate the available time to work for the line and Takt time for 100 units per day

Approach: You will need to calculate the minutes available in the working shift and then subtract the non working time

Solution: 8 hours shift is 480 minutes. Take away 30 minutes for lunch, 30 minutes for breaks (2x15 minutes) 10 minutes for a team briefing and 10 minutes for basic maintenance work;

The available time to work = 480 – 30 – 30 – 10 – 10 = 400 minutes

If output demand was for 100 units a day and you were running one shift, then the line would be allowed to spend a maximum of four minutes to make a product in order to be able to keep up with demand of 100 units per shift.

The **takt or cycle time** would be 4 minutes.

4 Capacity Planning

When capacity needs to be increased or decreased, the operation must consider how this is going to be achieved. This is a key decision as the organisation will have to make investment decisions based upon what level of capacity is to be selected and when it is to be provided. The operation has several ways in which it can respond to the changes in demand with its provision of capacity. The decision to provide capacity depends upon the selected strategy and the ability to store the product or timeliness of service production. The timing decisions of how and when to provide capacity need to be determined in line with demand.

4.1 Capacity Planning Methods

The organisation has 3 main choices;

1. It can provide capacity ahead of the forecast so that it is ready to respond immediately which is known as a **capacity leads demand** strategy.
2. It can provide capacity as demand changes so that it expands and contracts its capacity to follow demand, which is a **capacity matches demand** strategy.
3. It can wait to see what demand is and then respond after it is confirmed, a **capacity lags demand** strategy.

Capacity Leads Demand

It is possible to have capacity ready to react to an increase in demand as ready and available capacity. This is where a buffer is provided in order to allow the operation to react quickly to increases in demand. This strategy adds capacity in anticipation of extra demand and is therefore an opportunistic strategy with the purpose of attracting customers away from competitors. This capacity strategy has an advantage in that the operation is ready to satisfy customer demand and meet short term opportunities.

However there is a risk of demand not rising and the operation is then left with the wasted costs of unused capacity.

It is a more expensive way of providing capacity as it requires investment to be made ahead of demand, but it is a useful strategy if the organisation is trying to build market share and the benefit of establishing a customer relationship outweighs the cost of providing excess capacity.

An example of a capacity leads demand approach would be an extension to a lecture theatre being built before student numbers were confirmed.

Capacity Matches Demand

For the provision of capacity in line with demand then this strategy is adopted. This is done by adding capacity in measured amounts in response to changing demand in the market. This is usually accomplished by flexible addition of capacity either from flexible labour or flexible facilities that are able to meet the demand upon requirement. Either good planning is in place or there is a risk of underutilised resources.

This strategy relies heavily on forecasting and accurate information as investment decisions are made in line with the forecast. Incorrect forecasting will cause missed opportunities or wasted resources.

This often happens in services where staff are the flexible resource and can be brought in to cover peak demand yet sent home in quieter times, such as a toy store catering to Christmas demand or a restaurant expanding and contracting capacity in line with anticipated peaks and troughs in customer demand.

Capacity lags demand

Here increments of capacity are only added after the demand has increased by providing capacity after the demand rises. This allows the organisation to provide capacity with certainty and reduces the risk of incorrect investment into capacity increases. However this method does rely on the ability to provide products and services on short lead-time and assumes that the customer is prepared to wait.

This is less risky than providing investment ahead of demand; however, it has the disadvantage that customers may not be prepared to wait for the product or service and opportunities can therefore be lost.

Producing products on a lead time can be frustrating for customers, it can be almost impossible to buy a sofa from a store and have it delivered on the day, most have a four week lead time to allow the manufacturers to plan their capacity ahead of time. This is becoming an increasingly unusual strategy for consumer goods as consumers are often less tolerant of waiting.

4.2 Capacity Timing

The ability to increase or decrease capacity can be viewed in 3 time phases; short term, medium term and long term.

Short-term planning – this is a reactive time scale and can be as immediate as adjusting capacity on the same day or on a time scale of up to around 3 months (depending on the industry) Here, only flexible resources can be applied to increase the capacity. It may be costly to the operation as the speed of readjusting the resources may be higher on short term timescales. In many cases employees are the most readily available resource. Examples of this may involve measures such as;

- Over time for existing staff
- Having multi-skilled staff who can be reallocated to where a bottle neck has occurred. An example of this could be the tannoy call in a supermarket requesting, *'all till trained staff to report to the checkout'* in order to increase the capacity for payment, where queues are backing up at the checkout.

Medium-term planning – this time scale is beyond the immediate managing of the operation and has a horizon of around 3 – 18 months. This gives the operation more time to make plans to adjust capacity and therefore the changes are more significant than the short term plans.

- Hiring or firing contract staff
- Leasing in facilities, for example if processing calls, additional call centre support can be hired.

Long-term planning – this planning is a time scale beyond 12- 18 months. Here the investment decisions tend to be more significant and will link to the strategy of the operation. The changes will take a long time to implement but are also difficult to reverse. There are many more options available to consider with long term decisions relating to capacity and the possibilities for increases are far greater. They could include;

- New trained full time staff or fire existing staff
- New processes that may be faster
- New Machinery on a manufacturing line
- Information systems or technology can be applied to increase efficiency and capacity
- Additional facilities

With each of these options the ability to utilise the adjustment depends upon the nature of the individual process, a fast food restaurant may be able to hire and fire staff as training may take minutes for a new employee and investment is low, however a hospital may not be able to hire and fire skilled surgeons as they take many years to train and their skills maybe unique to the particular process they are involved in. Training methods for a car manufacturer rely on the skill of an individual

and are not easily replaced so a hire and fire policy is less effective, however if an operation were to invest in more technology that could replace many of these skills, then the staff would be more expendable and the flexibility of capacity would improve.

Short Case: The City of Manchester Stadium

The stadium used by Manchester City football club was originally designed to be the central arena for the Manchester bid to host the 1996 Olympics. When the games were awarded to Atlanta, the City of Manchester refocused their efforts on the Commonwealth Games bid for 2002, which they won.

The stadium was originally planned as an 80,000 seat arena for the Olympics. This was revised down to a 60,000 capacity stadium for the Commonwealth games. However, the Council's main concern was that the stadium should have a sustainable future so the plan was revised down again to accommodate a future for the stadium as the new ground for Manchester City football club to replace their Main Road Stadium in Moss Side. The revised plan meant that the capacity for the Games in 2002 was 38,000 which then rose to 48,000 in 2003 when it was handed over to Manchester City football club.

Construction of the stadium took 3,000 workers just over 2 years to complete and was handed over to the organisers four months ahead of the games. The Commonwealth Games was a spectacular success both for British athletes and the City of Manchester. However, there was no sentiment shown when the bulldozers moved in just hours after the closing ceremony. The track was removed, a third tier of seating was added and the central pitch was lowered. The conversion costs of £30 million were met by the football club.

This plan was not without criticism, as there were many calls by leading athletes for a large athletics stadium to be kept. However, the stadium from both Sydney and Atlanta Olympics became rugby and baseball grounds respectively. The Manchester Stadium has also been used as a concert venue and has a capacity for 60,000 fans, one of Europe's largest open air concert venues.

In 2010, an application was granted to expand the capacity to 60,000 for football fans. In 2002 the Manchester Commonwealth stadium had two tiers of seating, but after the stadium's conversion to a football ground, it had three tiers of seating. There are 2,000 parking spaces at the stadium itself with a further 8,000 spaces provided locally, there are 2 train stations within a half hour walk of the ground and for concerts and special events a bus service is set up. The stadium is used twice weekly during the football season, hosts conferences, major sporting events and even weddings.



Sources

<http://www.gameslegacy.co.uk/cgi-bin/index.cgi/30>
http://www.football-rumours.com/manchester_stadium.htm

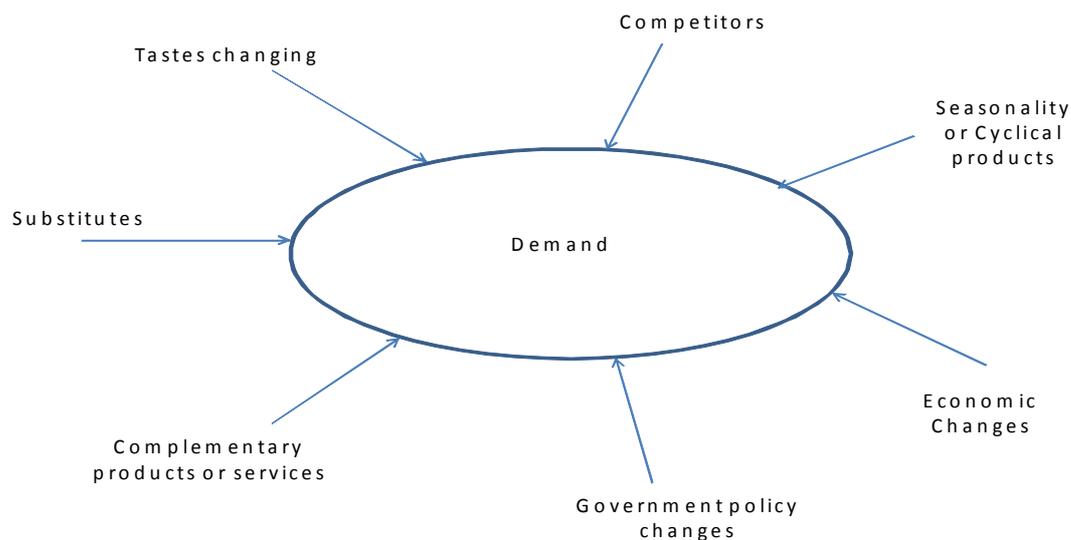
Questions

1. How has long-term planning affected the final result of the City of Manchester stadium?
2. Why was the original capacity for the Commonwealth Games stadium reduced from 60,000 to 38,000?
3. What other large facilities have changed use form their original purpose and have they been as successful as the City of Manchester Stadium?

5 Anticipating Changes in Demand

Demand can be volatile and is something that tends to happen 'to' an operation. It exists outside the organisation and is therefore difficult to control. Demand may change due to a number of factors or circumstances. As it is in the future and yet to happen and is subject to many influences;

Figure 7.3 this diagram shows the external factors affecting demand



- Tastes changing - these can be hard to plan for. Fashions change and new ideas occur. This can happen slowly overtime and be a natural trend towards an idea, for example the trend towards a more environmentally-friendly life style has been a slow but steady evolution and products and services have been able to follow this trend. Alternatively, they can be sudden swings or changes that are virtually unpredictable, for example an influential celebrity may 'adopt' a product or service and create a sudden increase in demand.
- Competition – the demand for an organisation's product or service will be highly dependent on the actions and reactions of its competitors. The product the organisation provides may have a steady demand but should a competitor bring out a new, improved or even cheaper version of the product then this will affect the demand horizon. It is important for the organisation to be aware of their competitors and the impact they may have upon future demand.
- Substitutes - this can be similar to the affect of competitors, mentioned above, as there may be an alternative product that will become available and divert the market demand for the product.

- Seasonality - many products are naturally seasonal. Ice cream, fireworks and fluffy Valentines-day teddy bears have a time period where demand will peak. This relates to the product or service use; how specific it is depends on how much it will be demanded at a given time. For example, sun screen sales in the UK peak in June and July, yet there will be a lower demand during the rest of the year for travel abroad, whereas Christmas cards only have the short window of demand in the months up to Christmas (generally August to December these days) and then demand virtually evaporates outside of this window. With a seasonal event, the previous year's demand can be used when predicting future demand patterns.
- Cyclical events are similar to seasonal demand peaks in the fact that they may be known events, but may not occur in line with a defined calendar time frame.
- Special events- there are one off unique events that have very little or no demand pattern or history. These are the most complicated to plan or predict demand for.
- The external environment may also affect demand patterns, tax and mortgage rates, government policy may also affect the demand for a product or service.
- Economic changes will have a significant impact on demand for goods and services and will affect buying habits and behaviours.

The key point is....

It is most important to try to understand what the demand pattern is likely to be and how reliable it is. The operation should then be aware of the potential impacts to their demand.

Short Case: Unpredictable demand

The BBC's Horizon program screened the results of a study of **Boots No7 Protect & Perfect Beauty Serum**. A team of dermatologists lead by Professor Chris Griffiths carried out an independent study into cosmetic anti-ageing products. The effect of No7 Protect & Perfect was compared against *retinoic-acid*, a prescription drug used to treat severe photo ageing of the skin, and the only product known to have a clinical effect on facial lines and wrinkles. The research revealed that in the laboratory tests pioneered by the Manchester team, No7 Protect & Perfect Beauty Serum really can repair the damage in photo - aged skin that is associated with fine wrinkles.



In the days following the airing of the programme the £16.75 Boots serum sold out. The sales volume – usually 1,000 pots of the lotion a week - reached 60,000 in the first 10 days. Ian Filby, The Beauty Director of Boots said: "We have been overwhelmed with the response following BBC Horizon, with women literally racing each other to get hold of the last product in stores. We are getting more stock out to our stores on an hourly basis to manage demand."

To cope with the sudden increase in demand, boots issued a one product per customer policy in order to dissuade consumers reselling product on auction sites. A waiting list was then introduced where approximately 50,000 customers signed up for the product; orders were being received from as far

afield as the USA and Australia. Despite the efforts of Boots to supervise sales the product was being sold on E-bay for up to five times its original price.

It took several months for supply to be fully restored. Despite other more critical studies the product continues to be a top seller for Boots and the range is sold around the world. Protect and perfect Beauty serum remains the fastest selling product in Boots history.

(Sources; Early rush for anti-ageing cream <http://news.bbc.co.uk/1/hi/uk/6623709.stm>, Finch, J. (2007) Serum to tackle age-old problem is here. Prepare for mayhem <http://www.guardian.co.uk/lifeandstyle/2007/may/03/shopping.science>, Poulter, S. (2007) The £17 cream even scientists say can banish wrinkles <http://www.dailymail.co.uk/femail/article-445488/The-17-cream-scientists-say-banish-wrinkles.html#ixzz0rlcaXcTA>)

Questions;

1. What action did Boots take to cope with this unexpected surge in demand?
2. What risks were there for Boots when undertaking this reaction to demand?
3. Think of other products or services where there has been an unexpected surge in demand. What have been the successes and failures of the operations dealing with these surges?

6 Forecasting

Forecasting is the prediction of future demand based upon qualitative or quantitative measures. It is an attempt to predict the future – which of course is an unknown. There are many different alternatives that a plan could take, and hence forecasting can sometimes be an educated guess.

The forecast is selected by the scheduler in the operation and needs to be based in the short, medium and long terms. Many subsequent decisions are based upon the forecast:

- **Material purchasing and expediting:** bought-in goods need to be ordered and allocated in line with the forecast. An incorrect decision may leave the operation with excess stock to store or with the costs of extra material to be expedited in.
- **The product mix:** what products or services are to be made and when are contained within the forecast. If this is incorrect the operation may be left with unsold finished goods and missed opportunities.
- **Staff levels and over time decisions:** staffing levels are based upon the forecast; recruitment decisions are made in line with medium and long term plans, over time may be required as a short term solution. Operations with a skilled staff base need to make any decisions to increase capacity well in advance of a demand increase, as it may take longer to recruit a specialist rather than an unskilled worker.
- **Capital investment decisions** will be made in line with forecast information. The decisions to expand or contract facilities are not without risk when based on a potentially uncertain future.

When accurate, forecasting, allows for timely planning of facilities and resources. When incorrect, it can result in missed opportunities or capacity wastage or extra costs in expanding the capacity. Organisations may heavily invest in information that may be relevant to future demand patterns or volumes

A forecast can provide a plan for the volume likely to be demanded over a time period, however this is not enough. It may be more useful to an operation is to understand those swings and movements in demand and timings and events that may affect demand. For example; past experience has shown the insurance industry that the majority of calls for membership are between 6pm and 9pm in the evening; hence they are able to adjust their capacity in their call centres accordingly. However more information is required. It may be useful for an insurance company to consider:

- Is a Monday likely to be busier than a Friday?

- Are there some weeks where membership is more in demand than others?
- Are there any significant events, such as a bank holiday that may alter demand?
- What additional and unpredictable events may occur that may change demand?

6.1 Forecasting Models and Techniques

Forecasting models can be used to help predict some of the answers to these questions.

By understanding how the forecasting information will be used, for example in work plans, investment or supply chain information, it is possible to be aware of how much needs to be invested in forecasting models. It is also important to know what timescale it is required for.

Questions that a manager should consider are;

- What is the time frame for the forecast - is the forecast for short/medium decision making or for long term planning?
- What decisions are to be made based on the forecast?
- Who will be using the forecast?
- What is the risk exposure to the organisation of an incorrect forecast?

A scheduler, who will be responsible for planning the production and capacity timing, will often use known or planned orders for the short term element of the forecast, but by the medium and long term time horizons the information is based upon gathered data. If a product is sold on a four week lead time then the first four weeks of a schedule is largely fixed to fulfilling these orders. However beyond this the operation needs to plan what it is going to make and this is where the scheduler will have to rely upon a forecast to estimate future demand.

Methods of forecasting can be based upon quantitative or qualitative methods. These approaches can be used independently, but are not mutually exclusive. Qualitative methods use appraisal, judgement and experience to estimate demand swings. These methods take into consideration a number of factors that may be based on past experiences, opinions or even hunches. Quantitative methods reply upon data based on factors such as previous sales, which identifies future trends. When combined these methods can be a very effective.

6.2 Qualitative Methods

The three main methods of qualitative analysis are;

Market Surveys

Here customers are surveyed for their future buying habits. From this data conclusions can be drawn regarding future demand.

Surveys are especially useful when looking at the potential for new products or services. This is because there is no previous information to be used as the product or service doesn't have a market or customer base so this level of concept testing and feedback can be useful.

The data collected needs to be from a truly representative sample of the market, which can be difficult to identify, plus the reliability of the consumer response can be questionable as the survey will ask for opinion, which may vary from person to person, Also a customer may tell the interviewers what they think they want to hear rather than their actual intended buying habits. Collecting the data and correct and detailed analysis of the responses can be time consuming and expensive.

For a new form of technology, such as a new feature for a mobile phone, market surveys can test the reaction of potential customers as to how they would use such a product feature and what would be valued about it. This can give an indication about how much of the product could sell and if the new feature would be considered to be significant enough to increase current volumes.

Delphi Methods

These are the detailed interviews and studies of the opinions of a panel of experts in the area. Individual opinions are collected anonymously and then circulated to other members of the group for interpretation and comment. This process is repeated until a group consensus is formed. This is a useful method as it evaluates different perspectives and provides a wider commentary on the collective opinion. The results from Delphi panels can be informative and provide useful insights into future demand. The method is likely to consider a wide range of ideas and intangible factors that may be missed in quantitative data.

The entire process can be very time consuming and relies on a good selection of experts, who can be difficult to identify. Sometimes their opinions do not reach a consensus and then the process is of less use. It can also be time consuming and expensive.

Scenario Planning

This is a 'what if' approach taken to the future demands and considers the potential scenarios an organisation might face and then analyses the demand pattern that may result from each scenario. A range of alternative outcomes are generated, and then the likelihood of each occurring and the risks involved are assessed.

This method can cover a wide range of possible futures however there will always be scenarios that are not considered and some of these may end up as significant.

An example of a scenario plan could be to consider the potential impact on demand for a building company if a General Election is due. The scenarios could be related to the relevant policies offered by the political parties related to funding housing projects should they win the election. This can give information to the organisation regarding the potential impact on their volumes.

Scenario planning works well where there is an established product or service where more variables and scenarios are understood. For new or emerging products there is too much that is unclear and unknown - for this method to be of much value.

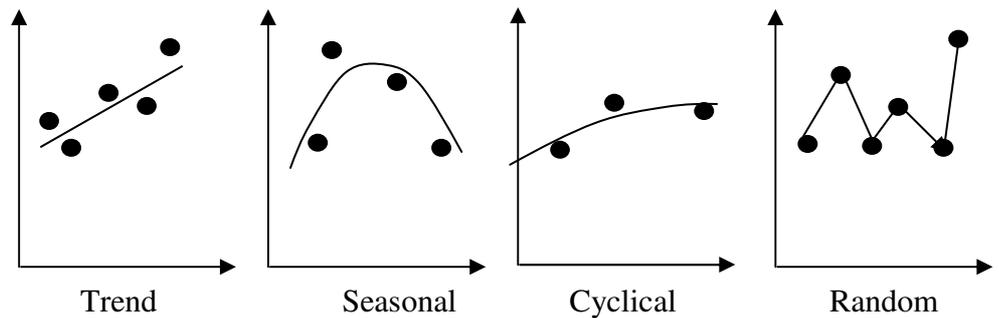
6.3 Quantitative Methods

6.3.1 Time Series analysis is a mathematical model that uses past data to predict future demand. It attempts to establish a pattern from previous demand time periods, and then this is put on a curve to be extrapolated into the future demand forecast.

Demand can be classified into four identifiable types:

- **Trend:** a gradual flow between the selected points which is carried on to the future. Factors contributing to trend may be the economic background which may be growing.
- **Seasonal demand:** a regular pattern year on year. When this data is combined with qualitative data these patterns can be simple to understand. Seasonality usually within a cycle could be affected by high and low season sales
- **Cyclical demand:** a deviation from the trend line that lasts for more than a year – it is the ebb and flow that is apparent. Cyclical demand may be affected by many factors including economic or social conditions, fashion or technology development.
- **Random variation:** an unexplained deviation from the trend line that follows no pattern or repetition. When all known causes for demand in a forecast that can be identified have been accounted for then the remaining unexplained variations are referred to as 'random' variations, error or noise. Predictable variations fall into one of the previous categories, but may be classed as random if considered a true one off event.

Figure 7.4 below illustrates demand pattern for each of the following elements: Average demand for the period trend, seasonal demand, cyclical demand and random.



However, if the time period is short only random variations or error are the concern of the forecaster, whereas, over longer horizons all four effects may be detectable.

6.3.2 Causal Relationship Forecasting

This is an interesting mathematical analysis where there is an attempt to link the demand pattern of a product or a service to the time line of an independent variable. It analyses a leading variable and observes a direct reaction in demand to changes in that variable.

Examples of this could include; the demand for umbrellas linked to rainfall or the demand for New York City breaks linked to the exchange rate between the dollar and the pound or euro.

The first step is to identify if the relationship is actually linked or if it is just coincidence. To discover this, the data needs to be analysed for the irregular peaks and troughs in demand.

Causal methods are used when there is historical data available and the relationship with the variable to be forecast (referred to as the **dependent variable**) and other key related variables (referred to as **independent variables**) is known.

For example the demand for baby essentials is closely linked to the birth rate.

- The forecast variable will be **sales of products**
- The independent variable will be **the birth rate** for which there should be reliable current and past data.

The relationship between the two variables is of interest because it is required in order to forecast demand based on this relationship. In the example included here, it could be assumed that the relationship is linear (close enough to a straight line). This approach is referred to as **Linear Regression**.

6.4 Other factors

A Forecast or prediction of the future can seldom be a single perfectly accurate estimate. However, a poor match between demand and capacity where demand far exceeds capacity or where capacity exceeds demand should be managed.

However, there are other factors that are taken into account;

- **The product life cycle.** Product demand typically follows four phases: Introduction-growth-maturity and decline.
 - For products in the first two phases of introduction and growth; longer forecasts are necessary e.g. sales of High Definition TV sets or Microsoft Windows 7 operating system.
 - Whereas products in the latter two phases of maturity or decline, requires shorter more specific forecasts usually of staffing levels, inventory levels or supply chain calculations.

- **Business cycles** also are taken into account. Market growth or recession provides a general backdrop for forecasting estimates and time periods.

When creating a forecast, each factor must be taken into carefully considered and evaluated for the impact expected on future demand.

7. Capacity Strategies

When an operation is planning how much capacity it needs, it must also think about how it plans to react to the demand it faces. The operation must be aware of the options available to satisfy demand. There are three general strategies that can be used in the medium term: They are:

- Level production
- Chase demand
- Demand management

These strategies are not mutually exclusive and most organisations use a mix of these three, however it is likely that one method will dominate the strategy. These strategies are not reactionary to small daily swings in demand but look at demand over a longer time frame. Such decisions require planning and investment.

7.2.1 Level Production

Level production largely ignores demand fluctuations and works on producing units efficiently and then either storing the finished goods or satisfying demand from stock. Basically this allows an operation to manufacture efficiently and optimise capacity irrespective of the demand. There are two key conditions for this to work;

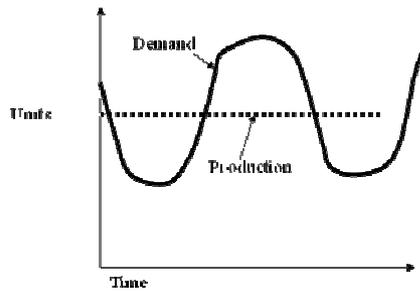
- The product must be suitable for storage, i.e. not highly customised and non perishable or with a reasonable shelf life.
- Demand must be relatively reliable – to avoid risk of large stock outs or excessive stock levels.

This level production strategy doesn't work for services.

This strategy forms part of a **cumulative plan**, where demand is satisfied over a period of time to suit the operation and stock holding and selling from stock allows production to be effectively managed by stock control allowing the operation to have the benefits of working to a less volatile plan.

Certain industries and products lend themselves to this strategy. For example, the demand for cigarettes is fairly consistent; there is a drop off in sales in January and February as a number of smokers make New Year resolutions to give up, yet demand returns, for those who fail to achieve this, to fairly consistent levels after this time and then peaks at around Christmas time for the party season. The product is straightforward to store and has a reasonable shelf life. As demand is fairly consistent the operation can feel confident about following such as strategy.

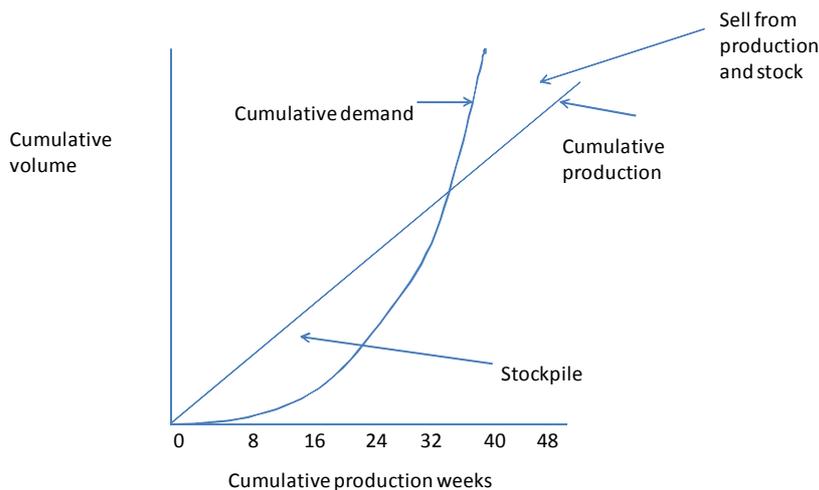
Figure 7.5 The level production strategy produces units to a consistent rate



Level Production

Aggregate planning is where forecasts are set to be as capacity-efficient as possible yet still aims to match the demand. This style of planning rises above the detail of the product or service mix and looks at capacity overall in line with the demand. It is a schedule that works in the medium term and makes decisions based upon staffing and stock holding levels or leasing decisions. This works as part of a level production strategy so that manufacturing is carried out to cover the demand so that stockpiling occurs ahead of selling from stock and not the other way round – it is hard to sell product from an empty warehouse!

Figure 7.6 The level production strategy considers cumulative production ahead of cumulative demand



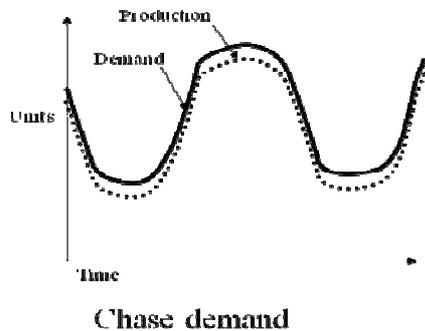
7.2.2 Chase Demand

This is where the operation attempts to follow demand by expanding and contracting capacity. There is always going to be an element of fixed capacity that can only be adjusted over the long term, such as a building or invested facilities, however other elements are flexible. By expanding and contracting these flexible resources the operation can minimise the costs involved with having excess capacity yet still increase capacity to meet increases in demand.

To follow such a strategy requires considerable planning in process design and staff training. The operation needs to consider the question *'if demand goes up and then falls, how are we going to expand and then contract our capacity without wasting money?'* In order to expand and contract capacity efficiently the process needs to be adaptable. This can be achieved by designing a very simple process so staff can be hired and fired or even multi skilled. Of course staff are not the only adjustable resource in the process – flexible machinery or the ability to make other products in times of low demand may help with this too – if a demand peak for one product could be combined with a demand slump in another then reallocating resources can allow the demand to be followed.

Both the level capacity and chase demand strategies follow a supply management approach where the provision of additional increments of capacity is adjusted. However the alternative approach is to manage or control demand.

Figure 7.7 A chase demand strategy has production following demand



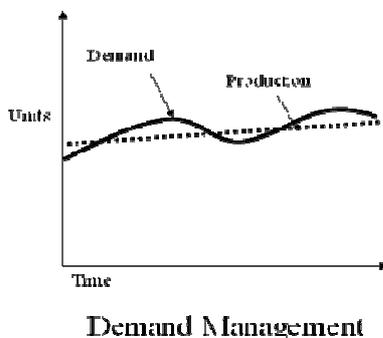
7.2.3 Demand Management

The attempt to manipulate or control demand may be unfamiliar to an operations manager, however it is possible by use of marketing tools to smooth demand in line with the required forecast for some products or services. Many of the tools suggested may be short term measures; however it is important that the adjustments made do not damage the potential long term demand. Some methods of demand adjustment may include;

- Varying the price. A discounted price for an elastic demand curve product can be a strong incentive for a consumer to buy more products. This happens regularly in supermarkets with price discounting, it works well if combined with additional advertising, mentioned below.
- Providing an incentive for off peak services or products. This works well with services to help smooth demand throughout a given time period. Good examples of this is an off-peak rate in the gym, or the 'happy hour' in bars, which are both used to flatten the peaks in demand and increase utilisation of wasted capacity at quieter times.
- Additional marketing can create awareness of a product that wasn't there before – this is most effective when combined with other methods such as price discounting.
- Alternative products or services could be provided using the same capacity and skills level but smoothing demand.

Demand management practices work in combination with the supply approaches mentioned above, it may not be possible to control demand to a level state, but it may help to smooth out the fluctuations of a chase demand strategy.

Figure 7.8 Demand management attempts to smooth demand fluctuations to allow a level production rate to be used



Contemporary Thinking: Yield Management

Yield management is widely regarded as an increasingly useful business tool and is generally used in operations where the medium term capacity is relatively fixed and there is no straightforward way to expand and contract the capacity provided. This is often the case in services with high investment facilities, variable costs are low, closed entry and exit points and fixed capacity prevail, such as hotels with fixed numbers of rooms or airlines with the fixed number of seats.

Yield or more commonly 'Revenue Management' was first introduced as the option to choose to maximise returns from business assets. Revenue management is founded in pure economics where the demand versus supply curves rule. This rule has been studied in business using computer software that looks at historical trends, fixed capacities and revenues generated and then 'forecasts the future', giving the business a 'set of rules' to apply to its assets in order to maximise their financial potential.

This strategy involves a set of strategies that an operation can employ to maximise revenue by either charging individual customers as much as they are prepared to pay or by extracting as much revenue from a process as possible. This asks the question; if a customer is prepared to pay more for a good or service, is it possible to charge them a higher price than another customer?

Certain conditions enable companies to use yield management. These include;

- Fixed capacity in the short and medium term.
- The ability to sell the service at different times to the different customers and in advance of the service being delivered.
- A market that has different and diverse customer requirements.
- A service that has some unique characteristics to avoid a homogenous market. This may be the destination or timing of a flight or the facilities or location of a hotel.

The 3 key strategies for maximising revenue through yield management are as follows;

1. Overbooking – this is where the operation books more customers than it can accommodate and presumes that there will be some 'no shows'. It has the rather obvious disadvantage of what will happen if all the customers do turn up! A policy of overbooking can only be effective if the take up rate for the service is predictable, and if the costs of compensating a customer if they are unable to have their booked service are not too high. The advantages for the operation include maximum revenue and full capacity. This has been a popular policy of many airlines, where flights are overbooked. Airlines have a large volume of data from previous flights to predict rates of no shows on various routes, customers who miss flights because of this policy are then compensated with alternative flights and possible upgrades. However recent legislation that requires airlines to financially compensate passengers as well as finding alternative flights and this extra cost has led to many of the budget airlines discontinuing this policy.
2. Price discounting – this is similar to a demand management policy to try to optimise the capacity at non peak times. In the case of price discounting the top price that can be charged is set for the time of highest demand and then reduced for less attractive time periods. A high peak time price may also be used to deter those booking at peak times to control demand. It helps smooth demand over a time period and maximises revenue. A good example of this is the holiday market – with a finite number of holidays available in Europe, peak time such as school holidays may incur heavy premiums yet there are many bargains to be had 'off peak'.
3. Varying the service type - this allows an operation to grade what service they are selling and charge different prices accordingly. This allows them to charge those who are prepared to pay more, the extra with some justification. Seats in a theatre will be grouped in order of quality and price, such as the dress circle seats being the best and most expensive and the stalls being the cheapest. Upgrading a service can differentiate it from the standard service and therefore higher prices can be charged. The advantage of this policy is that it is flexible in the short term; if an airline has over demand for economy seats for a flight but does not sell all its

business class seats, then a business class seat on a flight can be 'converted' to an economy seat to maximise revenue.

Stop and Think

Today, some argue that this practice is not so much Revenue Management as price manipulation, this being the price the business asks for the asset at a point in time. It may help the company increase revenues but is it always fair to customers who pay different rates depending on the time of purchase?

8. Queuing

8.1 Why are we waiting?

Several of the strategies mentioned above for the management of capacity, require the ability to be able to store products and sell out of stock in order to follow the demand line. In low demand times the product can be stockpiled yet in peak times the rise in demand can be satisfied from stock.

However sometimes it is not possible to build a product in advance, this is especially true of services, where by the very nature of a service the production and consumption are simultaneous. In these cases **queuing theory** is a more appropriate strategy. Queuing theory works on the basis that where the flow of consumers into a process is not regular then there will be an element of waiting involved where customers will have to queue.

This theory demonstrates that while some customers may be satisfied immediately, there will be periods of queuing for others as the arrival time of customers at the start of the process is not controlled by the operation.

The waiting line or queue effectively regulates the flow of customers into the process and turns the uneven demand into a manageable flow.

An example of this is a theme park; customers are free to walk around the theme park and take a trip on any of the rides on offer. The customers will arrive at the rides in an irregular fashion so there will be peaks and troughs in demand all day long. If a roller coaster can accommodate 100 passengers on a 20 minute ride and 600 customers arrive then they will be processed in 20 minute batches of 100 and the queue will be cleared in 2 hours. However the customers will arrive in varying groups and batch sizes in different timing this is known as the **arrival rate** of customers.

The customers entering into the process is called '**the calling population**'. The number of customers may be known and can be a known (finite) where for example a number of tickets have been sold in advance – an example of this maybe a theatre show, or can be completely unknown (infinite) where there is no advance warning of arrival number, an example of this maybe a post office.

When joining a queue, the customer is not usually engaged in the process but waiting to start the process, this leaves the customer free to either not join the queue, which is known as '**bulking**' or queue for a time and then leave the queue which is referred to as '**reneging**'. The operation itself also is able to prevent customers entering the queue, this may occur if there is too much demand for the service at the given time, this is where the customer is '**rejected**' by the queue.

The process servers

The points of sale or servers signify the start of the operation for most customers, they process each customer and then the customer finishes the stage of the process. There are several methods with using servers;

- There can be one main queue facilitated by several servers. An example of this could be the security checks at an airport; this has a central queue that filters off at the front to the next

available scanning machine. This can result in a very long queue but one that moves quite swiftly – the individual is not affected by a delay with another customer as they are able to be seen by the next available server.

- Each server can have their own queue and therefore there will be a number of queues side by side. This is common in supermarkets with the till checkout system. Here the customer has to select a queue and estimate the process time to find the shortest one. In the case the consumer has control over their selection but this can lead to frustrations for customers if other queues move faster than the one they are in.
- Lastly, there can be a sequential queue, this is where the customer joins a queue, is served and then moves onto the next part of the process in a continuing queue. An example of this could be in an accident and emergency department in a hospital where the customer (or patient) will join the queue to be seen by a nurse and then is moved to the second part of the queue to be seen by a doctor and then may move onto a queue for an x-ray. If all the treatments take the same time then the queue will flow evenly, if not a bottleneck will occur.

For a queue to be effective it must have clearly defined. Each customer needs to be aware of how they are expected to queue and in what sequence they are to be served. Customers boarding an aeroplane may expect to be seen on a **'first come first served'** basis but may also understand that there will be other passengers who are able to bypass the queue such as disabled passengers or travellers with young children.

Customer involvement in the process can be part of queuing theory. If a customer queues, they wait to join the process, however if the queue becomes part of the process the customer is technically served sooner. Using the queue to collect information such as pre orders in a cafe queue or collecting information relating to the process such as personal details and requirements engages the customer in the process and may reduce renegeing and frustration levels.

8.2 Customers perception of waiting in queues

The psychology of queuing has been a much discussed topic; operations need to know how their customers will react to a queue and if indeed having queues damages their operation in any way by customers switching to competitors who do not have queues or if it affects satisfaction and repeat purchasing. It is also important to understand how the disadvantages of queuing can be minimised by considering the feelings of the consumer towards the queue.

There are a number of common factors identified from a customer's perception about a waiting line;

- Time spent in a queue is seen as idle or wasted time. A consumer will make a judgement to trade-off the time spent in the queue with the perceived value of the service.
- Some consumers are prepared to pay a premium to avoid or reduce queuing time. This can clearly be seen at airport check-in desks where first and business class have their own server who does not support the main queue.
- Perception of the queue depends on the type of queue, and if the end result is worth waiting for. For example the queues for a theme park can be several hours but the ride is valued highly so there is a 'reward' for being in the queue.
- If the customer has information about the queue in relation to expected wait or position in the queue there is a much lower level of anxiety. Informing a customer of their situation empowers them to be able to make rational decision as to whether to remain in the queue or leave. Passing on this information often uses technology, for example automated call centre queues regularly inform customers of wait time and queue position.
- A customer queuing to start the process will feel more dissatisfaction than a customer who has already started the process; this can be seen in the accident and emergency queue where patients are seen by a nurse before they are seen by a doctor – therefore beginning the process more quickly but not necessarily shortening the wait time.
- Consumers have higher expectations for better service and queues are often seen as a negative thing.
- One of the lessons for Operations Managers is that customers dislike uncertainty. Waiting in a queue is an uncertainty.

Many things can affect happiness in the queue; it can be something as simple as good weather when queuing outside to the mood of the customer in the queue. If a customer is in a hurry and not anticipating a long queue this will have a more negative impact on their emotions than if a queue was anticipated.

It is important for an operation to be aware of and tackle frustration and irritation with the queue methods of achieving this involve information re timing and waiting times, position in the queue etc. Technologies have been devised to 'bust' queues; these may include hand held devices to take customers orders when in the queue, in this way customers are committed to the process and less likely to baulk or renege the queue. Many operations may take the customers' engagement in the process one step further and provide entertainment in the queue; this can be an effective policy for queues for music concerts, sporting events or theme parks.

Critical Perspective: Are all queues bad?

As a society consumers are less happy to wait for goods and services, as operations managers we have come to see waiting and queuing as a negative in a process that involves customers. Queuing has become so much part of the life of a consumer that it affects the way they make choices, plan purchases and allocate time to activities. But it has an important role to play in decision making;

- A consumer may be prepared to wait longer in order to get a lower price – they may therefore be happy to queue for lower price rather than pay a premium for a better service
- A queue can also be an opportunity to encourage impulse buys, which keep the customer shopping for longer than they had initially intended to do.
- Queues can also act as a signal or give information to a consumer. A night club with a queue may indicate that there is a popular party going on inside rather than a half empty club with no atmosphere.
- A queue may also be part of the anticipation for an experience such as the queue for a fairground ride can allow the customer to view the ride and see what they are about to experience themselves.

The key point is: to consider what the customer expects from the queue and aim to meet their expectations. A queue in itself isn't negative an unexpected, uniformed or particularly long queue is.

Short Case: Queuing for Wimbledon

As one of Britain's most popular sporting events, Wimbledon Tennis Championships attracts thousands of visitors a year; however it is one of the few major events to offer a number of tickets to spectators for sale on the day. There are around 500 premium tickets available for the 'show courts' and approximately 6000 tickets for the grounds. The number varies according to previous sales, the weather and the number of courts in play.

For a ground pass it is usually necessary to join the queue several hours before the grounds open. For a show court ticket many people choose to queue overnight to maximise their chances of getting a ticket for Centre, No.1 or No.2 Court. An area is created in Wimbledon Park for campers and facilities are provided. In the early morning campers are woken by stewards to pack up their kit and form a more regular queue.

There may be a wait of around 5 hours before the tickets then go on sale. Some members of the queue may be unsuccessful in their attempts and leave empty handed.

The queue at Wimbledon has become part of the Championships themselves and is known for being good natured and friendly.



http://aeltc2010.wimbledon.org/en_GB/about/tickets/queue.html (as adapted)

Questions

1. What makes a queue for Wimbledon different from the normal approach to queuing?
2. What are the positive and negative aspects for a customer queuing for Wimbledon?

9 Summary

One way to look at Capacity Planning and Management is that anticipating demand and investing in capacity is a fundamental part of operations management. Managing that capacity is what generates surplus or profit. How well we do this is a measure of our efficiency.

This chapter has introduced a number of approaches for dealing with everyday problems. These approaches should enable the anticipation of demand and the management of resources in the light of actual variable customer demand.

The chapter has presented:

- How capacity is measured and appreciates the dilemma faced by management in matching variable demand with variable capacity.
- The theory and practice the basic approaches to modelling demand patterns, arrival rates and queues; be able to use a number of approaches to forecast demand
- The principles outlined in the theory of constraints (TOC) and how constraints affect both throughput and profitability and help students carry out basic calculations.

Key Theories

Resource Mix

This is the collection of inputs or resources into a process that can be adjusted to increase or decrease capacity

Input and Output measures of capacity

It is possible to measure the capacity of an operation by analysing either inputs or outputs. Either method is correct; the selection of the method depends upon which measure is more logical. E.g. number of beds in a hospital is a logical input litres of oil for an oil rig is a logical output

Design capacity is a top level and unachievable result as it presumes 24/7 operation with no stoppages.

Effective capacity is the same as design capacity but subtracts stoppages such as lunch breaks and maintenance.

Actual capacity is more of a working number it looks what is actually produced and considers both planned and unplanned stoppages.

Theory of Constraints

This is a theory created by Goldratt which suggests that capacity can be almost infinite if constraints or bottlenecks are identified and removed. The theory states that once a constraint is removed capacity will increase and another constraint will form.

Utilisation and Efficiency of capacity

These two definitions are ratios of how well an operation is using its assets and resources to generate output. Efficiency looks at how much is produced consider what the operation would hope to make considering planned stoppages. This allows the operation to establish how much unplanned problems

are affecting the output. Utilisation looks at output compared to design capacity and demonstrates how well the processes are working.

Takt Time (or cycle time)

This is the maximum time allocated to complete a process. If the process overruns this time output will be reduced.

Level Production

Level production this is where the operation produces to a consistent level, unsold product is stored and when demand peaks it is satisfied from stock.

Chase Demand

Chase Demand is where the operation expands and contracts its to follow demand changes.

Demand Management

Demand Management is where the operation attempts to manipulate demand to smooth the peaks and troughs. An operation may use a combination of all or some of these strategies as they are not mutually exclusive.

Yield Management or revenue management allows operations to maximise revenue if they have fixed capacity, by varying price from customer to customer depending on their willingness to pay. Higher paying customers might receive additional features or benefits.

Key Equations list

$$\text{Capacity} = \frac{\text{Time available}}{\text{Time of task}}$$

$$\text{Efficiency} = \frac{\text{actual output}}{\text{effective capacity}}$$

$$\text{Utilisation} = \frac{\text{actual output}}{\text{design capacity}}$$

$$T = \frac{T_a}{T_d}$$

Suggested answers to stop and think question

Stop and Think

Today, some argue that this practice is not so much Revenue Management as price manipulation, this being the price the business asks for the asset at a point in time. It may help the company increase revenues but is it always fair to customers who pay different rates depending on the time of purchase?

To answer this question, consider what the advantages are to the customer who pays the higher price. A customer may benefit from a better view in the theatre or a more comfortable seat on a plane. However also consider the advantages to the consumer paying the lowest price; they may not be able

to afford the service if there was no differentiation in pricing as the average price would most probably be higher.

Review Questions

1. What would be the most effective measure of capacity for;
 - a. A car manufacturer
 - b. A music concert
 - c. An airport

3. How could the capacity of a coffee shop be;
 - a. Increased in the short, medium and long term?
 - b. Decreased in the short, medium and long term?

4. When forecasting demand, what qualitative methods can be used? What are the advantages and disadvantages of these methods?

5. When an operation has a seasonal product, what strategies are available to them when planning capacity?

6. What must an operation consider when using historical data to predict future trends?

7. How can queues be managed effectively to reduce waiting time and customer frustration?

8. Using the Theory of Constraints, what are the most common constraints in a;
 - a. Car manufacturing plant
 - b. Post office
 - c. Dentists

Discussion Questions

1. How can having multi-skilled staff increase the capacity of a hospital? What Factors must be taken into consideration when increasing or decreasing hospitals capacity?

2. If an operation adopts a level production strategy, what features must the product have for it to be successful? When may other strategies be more appropriate?

3. How might an operation manage demand in a restaurant?

Problems

1. Calculating capacity measurements

A computer firm has a group of 50 computer consultants. These individuals visit firms in the area on or pre arranged visits or are called in for emergency repairs.

A call out fee is charged that covers the first hour of their visit. Beyond the first hour they charge in minimum blocks of 30 minutes. The average call out is 2 hours long.

The working day is usually eight hours long but allows 2 breaks of 15 minutes each and a half hour lunch break, leaving a 7 hour day. If holidays and illness are accounted for at 25% the 7 hours per day is actually $5\frac{1}{4}$ hour day

If actual work is only 500 hours billed in the week then

(a) What is the capacity utilisation of the team?

(b) What is their efficiency?

2. Calculating Takt time

A student registration line has a working day of 6 hours. There are 25 members of staff working during induction week. These employees have a half hour lunch break and two fifteen minute tea breaks each day. At the start of each working day, the process takes 15 minutes to set up.

Calculate the total time available to work for the registration team and the time that can be spent per student application (Takt time) for 2,000 students per day per day

References:

Aksin, Z., Armony, M. and Mehrotra, V. (2007), The Modern Call Center: A Multi-Disciplinary Perspective on Operations Management Research. *Production and Operations Management*, 16: 665–688

Goldratt El. M. (1999) *Theory of Constraints*, Published by North River Press

Heizer. J and Render, B. (2008) *Operations Management 9th edition* Published by Prentice Hall

Krajewski, Lee J. & Ritzman, Larry P. (2005) *Operations Management: Processes and Value Chains* New Jersey: Prentice Hall

“Future of revenue management — From the plane to the shelf”, by J. Bain, J. Journal of Revenue and Pricing Management Vol. 7, 3 302–306 & 2008 Published by Palgrave McMillan

Further Reading:

Pinedo, M (2009) *Planning and Scheduling in Manufacturing and Services 2nd Edition* Published by Springer

Key Terms

Capacity is the maximum goods or services that an operation can produce

Demand is the quantity of goods and services demanded by consumers

Scheduling is the matching the capacity of the operation with the timing of the demand forecast

Constraint is a restriction on the output of an operation.

Design Capacity is the expected output of an operation where there are no stoppages

Effective Capacity is the expected output of an operation considering planned stoppages for maintenance, shift change over etc

Actual Capacity is the expected output of an operation considering both planned stoppages and unplanned stoppages.

Efficiency is the output shown as a percentage of the effective capacity of the operation. It demonstrates how well the operation is working to expectations.

Utilisation is the output shown as a percentage design capacity of the operation. It demonstrates how well the resources are working.

Aggregate Planning takes a top level look at demand for the operation as a whole over the medium term. It does not differentiate between different products or services but looks at demand as a whole.

Level production this is where the operation produces to a consistent level over the medium term, unsold product is stored and when demand peaks it is satisfied from stock

Chase Demand is where the operation expands and contracts its capacity in the medium terms to follow demand changes.

Demand Management is where the operation attempts to manipulate demand to smooth the peaks and troughs. It is used in conjunction with the marketing department

Yield Management is a group of methods that assist an operation with fixed capacity to maximise its revenue and utilise the capacity it has to the operations best advantage

Process Servers are the points of sale or stations at the front of the queue where the customer is served.

Solutions to problems

Calculating capacity measurements

A computer firm has a group of 50 computer consultants. These individuals visit firms in the area on or pre arranged visits or are called in for emergency repairs.

A call out fee is charged that covers the first hour of their visit. Beyond the first hour they charge in minimum blocks of 30 minutes. The average call out is 2 hours long.

The working day is usually eight hours long but allows 2 breaks of 15 minutes each and a half hour lunch break, leaving a 7 hour day. If holidays and illness are accounted for at 25% the 7 hours per day is actually 5 ¼ hour day

If actual work is only 500 hours billed in the week then

(a) What is the capacity utilisation of the team?

(b) What is their efficiency?

Solution:

$$\text{Design Capacity} = \frac{50 \text{ workers} \times 7 \text{ hours per day}}{2 \text{ hour per customer call out}} \times 5 \text{ days of a working week}$$

$$\text{Design Capacity} = 875 \text{ customers per week}$$

$$\text{Effective Capacity} = \frac{50 \text{ worker} \times 5.25 \text{ hour day}}{2 \text{ hour per customer call out}} \times 5 \text{ days of a working week}$$

$$\text{Effective Capacity} = 656 \text{ customers per week}$$

$$\text{Efficiency} = \frac{\text{actual output } 500}{\text{effective capacity } 656}$$

Then the efficiency of the process is 76%

$$\text{Utilisation} = \frac{\text{actual output } 500}{\text{design capacity } 875}$$

The Utilisation of the process is 57%

Calculating Takt time

A student registration line has a working day of 6 hours. There are 25 members of staff working during induction week. These employees have a half hour lunch break and two fifteen minute tea breaks each day. At the start of each working day, the process takes 15 minutes to set up.

Calculate the total time available to work for the registration team and the time than can be spent per student application (Takt time) for 2,000 students per day per day

Solution: 6 hours shift is 360 minutes. Take away 30 minutes for lunch, 30 minutes for breaks (2x15 minutes) 20 minutes for set up;

The available time to work = (360– 30 – 30 – 20) x25 = 7,000 minutes

If 2000 students were to be processed a day, the maximum amount of time that could be spent per student application would be 3 ½ minutes.

Therefore the **takt time** is 3 ½ minutes