

**SWP 56/91 "CAPACITY MANAGEMENT IN SERVICES
AND THE INFLUENCE ON QUALITY
AND PRODUCTIVITY PERFORMANCE"**

**DR COLIN ARMISTEAD AND GRAHAM CLARK
Operations Management Group
Cranfield School of Management
Cranfield Institute of Technology
Cranfield
Bedford MK43 0AL
(Tel: 0234 751122)**



CAPACITY MANAGEMENT IN SERVICES AND THE INFLUENCE ON QUALITY AND PRODUCTIVITY PERFORMANCE.

Colin G Armistead and Graham Clark

Cranfield School of Management

Abstract

Capacity management in services to match supply and demand has a direct influence on the ability of the service delivery system to achieve service quality and resource productivity targets. The paper examines some of the propositions for influences of capacity management on quality and resource productivity and for managing capacity. In addition to the "Chase" and "Level" strategies for managing capacity in services suggested by Sasser (1976) a "Coping" strategy for capacity management is described which aims to improve the overall delivery of service quality to resource productivity targets.

Introduction

The literature on quality management in services has been expanding rapidly (Gronroos 1984, Collier, 1987, Berry et al 1990). However these papers more often than not deal with questions relating to the definition of quality and the identification of measuring systems for quality management. While they recognise the influence on quality of the perceptions of management and service providers they do not address in detail issues relating to the overall management of resources for the successful delivery of services. It is our contention that the aspect of the delivery of services which has received insufficient recognition is the management of capacity. An operations manager in a service business will either succeed or fail in the process of balancing quality of service and resource management, expressed in terms of resource productivity, depending on their skill in managing capacity to match demand. This paper examines more closely the factors in the management of capacity which affect quality and productivity performance.

There is an interaction between capacity management, quality management, and resources productivity or efficiency management which is at the heart of the planning and control process for operations management in services. A number of authors have identified some of the issues confronting operations managers in managing supply and demand in services which affect their ability to maintain quality standards while achieving productivity targets (Lovelock 1984, Rhyne, Heskett et al 1990). The main issue is dealing with changing levels of demand which can occur rapidly and with an element of uncertainty while having some limiting factors on the ability to alter capacity.

Capacity management in service operations is a testing activity for operations managers because the nature of the service delivery process and the involvement of the customers in the process restricts the normal options open for controlling the process to match supply with demand; namely, altering the capacity, holding and inventory in anticipation of demand, and requiring customers to wait for the service. There is

Dr. Colin G. Armistead and Mr. G. Clark, Cranfield School of Management, Cranfield Institute of Technology, Cranfield, Bedford MK43 0AL. Tel. No. 0234 751122. Fax No. 0234 751806

Paper to be given at the British Academy of Management Conference, Bath, September 1991.

not the possibility of producing the complete service package in advance of demand and holding it as an inventory. This real time element of service production makes the matching of supply and demand very important, particularly in capacity constrained services like airlines and hotels when the profitability of the operation is closely linked to the use of the current capacity.

Capacity management is the ability to balance demand from customers and the ability of the service delivery system to satisfy the demand. This places an emphasis on understanding first the nature of demand by forecasting (Lovelock, 1984) and second the options for managing capacity to meet the expected demand. Sasser (1976) has suggested two basic strategies for managing capacity in services of "Level" and "Chase", the former applicable where capacity is limited and hence the focus is on influencing demand to be in line with capacity, and the latter strategy being possible when supply can be changed to keep in line with demand. Consequently operations managers must understand the composition of their capacity, the degree to which it can be changed, and the speed of reaction (Slack, 1987), and the costs involved (Heskett et al 1990).

However it may not always be possible to match capacity with demand finely and in these circumstances a coping mechanism may be employed which can alter the nature of the service delivered from the original design and hence the customers' perceptions of the service quality. Operations managers must be fully aware of the implications of possible changes as must those carrying out customer satisfaction surveys if correct inferences are to be drawn from the data.

Understanding Capacity

Defining Capacity

The definition of capacity is not always easy for a service operation. What is the capacity of a lecture room, a surgical ward, a supermarket, or a passenger ferry? Is it the number of seats, the number of beds, the square meterage, the number of seats and berths? Whereas these measures may indicate key resources to be managed they are static measures and are therefore unlikely to take account of all the factors which affect capacity. Consequently a definition of capacity as *the ability to work off an existing demand* makes for a more dynamic measure.

Once capacity is expressed in these terms it becomes apparent that it is related to an output *quantity* from the service delivery system. Consequently capacity has a time dimension and is influenced by all input elements to the system. So a variable mix of product and service packages will give rise to variations in the input and demand placed on the service delivery system. This variation will, in turn, mean the capacity of the system will alter as the balance of the resources required alters with changes in the product produced.

Capacity Levels

Capacity also needs to be related to the level within the service company and here there may be up to four levels which are useful for complex service organisations, characterised by the network and branch structure. The four levels relate to different orders of scale and managerial responsibility:

- * Strategic Business Unit Network.

The capacity of the business unit network will depend on the extent of the network, the number and size of the branches or outlets. For some organisations, this level will also include the regional management structure. At this level it is likely that capacity will be defined in output per accounting period or per calendar year.

* The Branch or Local Outlet

The capacity of a branch will depend on its size and/or the degree of support it needs or gets from a centralised support operation. The branch is the smallest unit capable of delivering most of the services. Capacity is often defined in daily or hourly terms.

* The Team

Within most service delivery systems there are teams who are responsible for the provision of one discrete services in a multi-service offering from the branch (eg a kitchen team in a hotel), or one stage of the total service (eg the X-ray department in a hospital). As with the branch, capacity will be defined in daily or hourly terms.

* The Individual Resource

The individual resource may be one person, a person acting with equipment like a VDU, or a piece of equipment the customer uses. Operations managers must be aware of the capacity of those individual resources which restrict the output of the overall operation.

Frontline versus Support Activities

Service delivery in most services contains two components; frontline where there is direct contact between the customers and the frontline staff, and support where work is carried out as part of the service package without the direct involvement of the customer. The two types of activities may require close linking of their respective capacities as with the supply of information to counter staff in a bank or building society branch. Alternatively the two types may be very separate as with a cheque sorting facility which can have its capacity determined independently. The degree of close linking is influenced by the extent of customisation of the service package and the extent of customer participation. Separating the "technical core" of support away from the disruption of direct contact with the customer may increase the output at least for standardised services.

Process Stage Capacities

When service delivery consists of a number of stages the capacity at the different stages should ideally be in balance as for any linked productive system. When systems are simple this is not too difficult to manage. However as the variety of the task at each stage increases it becomes more difficult to maintain the whole in balance. It is often important to understand the balance of capacity at different stages in overall service delivery and the way in which quality can be affected. For instance in a customer service and support operation, if capacity is given to satisfying routine maintenance as it becomes due the level of demand subsequently on the operation due to failure is minimised. If capacity is not given to the routine maintenance demand rises and also perceived service quality suffers.

Understanding Capacity Range and Response

Capacity at all the levels will not be constant even for constant demand. The variability in the execution of work by people and equipment causes fluctuations. The fluctuations may be smoothed over a relatively long timescale so as to be almost constant as for annual capacity at the network level. However at the team or individual resource level the fluctuation on an hourly to daily basis may be such as to give difficulties in meeting demand.

The ability to change capacity to cope with changes in demand raises questions of the flexibility of the capacity (Slack ?). The important issues here are of range and response; how far can the capacity be altered? (can it be doubled or halved?), what is the effect on costing of changing capacity? and how easy is it to make that change? are there different cost and quality implications for the level of response? The question of range and response need to be examined in relation to capacity for the various levels.

In coping with the variability of capacity, the idea of *effective* and *potential* capacity is useful (Lockyer et al. ?). *Effective* capacity can be stated as the capacity which is immediately available to the operations

manager, whereas *potential* capacity is the capacity which could be made available if the operations manager has time to make arrangements to increase capacity in the near future. Both these capacity decisions are short term decisions and relate more to the team and individual resource level rather than to the branch or network which are associated with longer term increases or decreases in capacity such as building a new facility or purchasing additional equipment and recruiting substantially more personnel resource.

Factors Influencing Capacity

When deciding on the effective capacity it is necessary to consider the quantity and nature of the input factors and also the amount of work which needs to be done to produce the service. Here it is necessary to recognise the three adding value activities for all productive systems, namely, improving, transferring, and caretaking (Armistead 1990) The inputs to the service delivery system which are recipients of the added value are materials, information or people, as customer, clients etc. When considering *effective* capacity there are three main influences which need to be understood which relate to the work on the adding value activities :

- * First, the potential of the service delivery system to work on the main input elements to which value is being added (ie materials, people, or information). In order for a service delivery system to have the capability to carry out work it must have all of the necessary resources needed, made up of a mixture of materials, information, equipment, and service providers. Should any resource be absent the system will have zero capacity.
- * Second for there to be an real output, rather than a potential to work off load, there must be a supply of the factor to which value is being added ie people, as customers, clients, passengers, materials, for distribution, or information, as for broadcasting.
- * Third the condition of the input elements to which value is being added will also affect the capacity. For example the heterogeneity of customers in their skills at taking part in the service delivery process, the variation in the accuracy of information to be processed, or the range and suitability of materials for the delivery process.

These three aspects of capacity restate one of the basic rules of capacity that variety in the inputs and/or in the demanded outputs from a service delivery system reduces effective capacity.

What Limits Capacity

Even though the resource which limits capacity may often be evident within a service delivery system, it is worthwhile from an analytical standpoint to examine in detail the make-up of the resources which compose the capacity of the service delivery system. The analytical approach is the following:

1. Establish the most appropriate unit for the statement of capacity.
2. List the main elements which make up capacity; included will be statements relating to all or some of the following resources, people, information, facilities, equipment, and materials.
3. List the factors which influence each of the main input elements contributing to the capacity. For example the information element may be used to record time and frequency factors including relevant service times.
4. Identify the resource which is the limiting resource causing bottlenecks and preventing the fluid change in effective capacity.

This process can be followed for both *effective* and *potential* capacity to give an understanding of the steps which might be taken to alter the capacity and the decisions which might be made. The process gives a better understanding of the *range* and *response* aspects of the capacity. It may be seen that there will be difficulties in altering the *potential* capacity because of the problems or expense of gaining a limiting

resource or that the changes in capacity will involve large step changes which need to be supported by expected increases in demand.

Understanding Demand

Capacity management is concerned with balancing the ability to produce added value work and the demand for that work. The above discussion has dealt primary with the provision of the capability. The other side of the equation is knowing what is needed and when. This information is contained in a knowledge of the volume, variety, and variation in demand and in the nature of that demand (Armistead et al ?).

Volume is a statement of the trend in demand over a period of time, typically between one month and a year. Variety gives information on the range of services being offered. Variation relates to changes in the volume demand around the general trend for the volume and also variation caused by the heterogeneity of customers.

The nature of demand and variety of services and the volume are factors which should be established at business unit network level rather than operational level. The levels of demand for services may be estimated by the application of forecasting techniques or by known demand from existing customers orders (although even in these cases there are elements of uncertainty in many services caused by the no-show syndrome). While historical trends are helpful in forecasting demand they do not take account of random effects which can influence demand in the short term, the effect of changes in the weather can be very significant.

Operations managers must be able to translate the demand into some expression of work content, ie the amount of work which is required to produce the service. The statement of work content may be expressed in terms of attendance hours on the basis of passed experience or by application of work measurement techniques. When the tasks to be performed are simple and repetitive it is relatively easy to make the calculation. However for more complex tasks often including diagnosis the prediction of the work content of the service time is more difficult to know in advance and may be dealt with by planning a partial allocation of capacity knowing that subsequent work will need to be done to complete the service delivery. Such is the approach often taken with professional services where an initial consultation is given at a set work content or service time and further work is deferred and completed in subsequent steps at a latter time. This enables the demand for capacity to be established more accurately as the process proceeds.

Capacity Management - Creating the Balance

There are two polar opposites for managing service capacity, one to hold capacity steady while influencing demand and the other of changing capacity to stay in line with demand (Sasser 1976). In reality most service providers use a mixture, although in the case of service delivery systems which have a clear capacity constraint there is a bias towards level capacity; this is evident at the branch level for hotels and airlines.

The "Chase" strategy described by Sasser entails controlling the level of capacity by changing the extent of resources by:

- * Altering the number of service providers and/or the hours worked, often involving the use of part time staff.
- * Sharing capacity between different parts of the service delivery system.
- * Transferring resources typically between back room and front office.
- * Using outside suppliers through subcontracting or leasing to provide resources.
- * Asking customers to provide more resource by way of self service.

The aim in the *chase* strategy is to maintain capacity closely in line with effective capacity thereby ensuring maximum efficiency and attainment of service quality levels.

The "Level" capacity strategy recognises real constraints in altering effective capacity and seeks to influence the level of demand by way of:

- * Price changes.
- * Advertising and promotion.
- * Developing off peak demand.
- * Use of appointment and reservation systems.
- * Making customers queue for the service.

Level capacity strategies are increasingly linked to yield management systems where effective capacity is constrained, for example in airlines, hotels, and car rental. In yield management the aim is not necessarily to gain the highest utilisation from the limited resource but rather to maximise the revenue yield from pricing differentials. This brings with it an additional control mechanism of "overbooking" which may result in customers not being served at the time they expect.

The Link between The Management of Capacity, Quality and Resource Productivity

Operations managers use capacity management to minimise the trade-off between resource productivity and quality. Quality and resource productivity are important in the strategy context as they impinge on the ability of the service organisation to attain its competitive strategy described by a combination of perceived added value (by the customer) and price (Bowman 1990). Service quality and customer satisfaction is aligned to perceived added value for the customer and resource productivity and unit costs affect prices and profitability.

The emphasis towards either quality or efficiency will to a large extent be driven by the competitive position of the service company. If competitive advantage is gained through service quality at a high price (compared to competitors) there will be more latitude in the trade off resource productivity in this pursuit and hence a tendency to indulge in capacity redundancy at times. If the service business is competing more on price the resource productivity is more likely to take precedence over quality, which may be allowed to fall in the area of absolute quality and customer satisfaction. However with increasing customer expectations of service quality service organisations are being forced to maintain higher levels of service at lower prices in times of recession and price wars than perhaps has been the case when demand has been higher than supply.

Also many services, notably in banking, hotels, and airlines, are no longer gaining clear competitive advantage from the provision of certain levels of quality as these are now regarded as a norm for the industry to which all providers in a competitive group must aspire. So reducing quality as part of a trade-off against resource productivity is less well definite. A much greater sophistication is required which can allow the less visible aspects of quality to be reduced, thereby keeping the whole service delivery process under far greater control. This in turn requires an understanding of the capability of a given level of capacity to deliver different dimensions of quality. The trick, it would seem, is to make sure that capacity is provided to support the critical dimensions of service quality. The critical dimensions are usually the dimensions which win customers or those which if they deteriorate too much lead to loss of customers (Armistead 1990).

The elements of resource productivity which are important are resource utilisation, input costs, and efficiency (related to doing things right first time and throughput time). In a similar fashion the range of elements which make up quality management may be summarised for present purposes into four categories, absolute perceived quality, "right first time", holding to specification, and customer satisfaction.

Poor service providers are likely to be often out of control of their capacity and either fail to satisfy demand or maintain excessive inefficiencies. Good service providers will struggle to deal with the two extremes of when a *chase* strategy runs out of capacity and becomes *level* and when demand drops away from effective capacity so that resource productivity drops.

An indication of the changes in demand and the consequential problems comes from a survey aimed at understanding the management of productivity in services (Armistead 1991). Demand for services had increased in the previous 5 years for 84% of the respondents from across the services sectors along with an increase in the number of services which any one organisation was delivering, ie an increase in variety. Only 13% of respondents indicated no problem in managing capacity to meet demand and daily and monthly variations presented the greatest problems for 32% and 20 % respectively caused by uncertainty in the demand and difficulties in responding rapidly with changes in resources. The same survey found that frequent changes in demand contributed towards a failure to improve productivity in 15% of the cases.

It is in this context that dealing managing capacity to match demand becomes more difficult and additional capacity strategies are needed: one to deal with running out of capacity and the other to deal with the failure to influence demand.

The "Coping" Strategy.

As service operations managers become cleverer at managing capacity and balancing it with demand it is at these capacity break-points where things start to go wrong. What is needed is some control over the drop in service quality which almost inevitable at these times or an accurate costing of steps to create new capacity and of the decision making process. It is clear that in addition to the two strategies proposed by Sasser (1976) we need an additional *coping* strategy which applies to the short term inability to match effective capacity and demand. The *coping* strategy is appropriate for those circumstances which can be characterised by "being busy" or "being slack".

As a precursor to developing a *coping* strategy it is obvious that any service provider should fine tune their own combination of the *chase* and *level* strategies by:

- * Improving their forecasting capabilities
- * Setting clear service quality targets
- * Setting clear resource productivity targets
- * Understanding critical and hygiene dimensions of their service quality (see Armistead and Clark 1991)
- * Understand the possible failure points in service delivery (see Armistead and Clark 1991)
- * Understand the bottlenecks in the service delivery system (see Armistead and Clark 1991).

This approach has been taken by a range of service providers including a bank which uses forecasts and a detailed load and capacity plan to control its operations, a holiday leisure complex which manages demand and capacity across a range of activities for up to 3500 holiday makers at one time, and a charge card authorisation centre which maintains a 3 second telephone answering target for almost 100% of the time.

There are then four areas when the *coping* strategy becomes necessary:

1. When a *chase* strategy becomes *level* in the short term because effective capacity cannot be increased to meet demand, usually because this would lead to under utilisation of resources.
2. When a *chase* strategy becomes *level* because it is not possible to reduce the level of resources any lower in the short term.
3. When a *level* strategy fails to stifle demand or overbooks.
4. When a *level* strategy is unsuccessful at filling effective capacity.

There are a number of possible actions possible in the four coping cases:

Coping: Chase Strategy with High Demand

At risk here is a drop in perceived quality of the service for the service receivers. The actions which can improve the situation are:

- * Understand what suffers for the customer, for example front line staff are less attentive, or service times extend.
- * Learn to recognise when the limiting stage for capacity is reached. Most commonly queues will begin to develop.
- * Limit the service provision, for example limited menus, and encourage self service.
- * Encourage subsequent complaints from dissatisfied customers to gain the opportunity to recover in the eyes of the service receiver.
- * Build expectation of lower service provision by warning customers before hand.

Coping: Chase Strategy with Low Demand

In circumstances of low demand but without the ability to reduce resources any more, resource productivity targets are at risk. The possible actions are:

- * Accept the time as rest and recovery for front line providers to combat "burnout".
- * Train service providers
- * Complete other tasks eg cleaning, maintenance.
- * Finish previous tasks which have been left unfinished in busy times.
- * Improvement teams for quality and efficiency.
- * Plan reduction in provision so as not to affect service quality. For example sequential closure of parts of a restaurant to maintain the atmosphere for the reduced number of customers.

Coping: Level Strategy with High Demand

In circumstances when it has not been possible to limit demand to match effective capacity or where booking systems like yield management systems have encouraged overbooking the following actions are possible:

- * Warn customers by way of notices or media announcements.
- * Establish recovery procedures to deal with overbooking

Coping: Level Low Demand

In circumstances when it has not been possible to stimulate sufficient demand to match effective capacity.

- * Improve service provision eg increase individual space on an aircraft.
- * Transfer service providers to other maintenance tasks.
- * Train service providers.

In all of the four situations it would appear to be good practise for operations managers in services to be increasingly aware when the limits to matching demand and effective capacity are being reached, and for them to take action at these times rather than ignoring what is happening.

Conclusion

Managing capacity to match demand is at the heart of the operations management task, and success or failure directly influences the attainment of a competitive strategy delivering a mix of perceived added value (ie service satisfaction) and price. The *chase* and *level* strategies of Sasser are in themselves inadequate to handle the limits of effective capacity. The *coping* strategy has been presented as the additional tool on the basis of observations of operation managers in service organisations. We are currently engaged in a larger research project to establish how a larger sample of service organisations manage capacity and demand and the effect on service quality and resource productivity.

References

- Armistead, C.G., Clark, G.R., 1991, "Service Delivery", Management Service Quality, July 1991
- Armistead, C.G., 1991, "Resource Productivity Management in the Services Sector", Cranfield School of Management Working Paper 20/91
- Armistead, C.G., 1990, "Service Operations Strategy: Framework for Matching the Service Operations Task and the Service Delivery System" International Journal of Service Industries Management, Vol, No
- Armistead, C.G., Slack, N, Johnston, R., 1988, "The Strategic Determinants of Service Productivity", International Journal of Production and Operations Management, Vol. 8 No. 3.
- Berry, L.L., Zeithaml V.A., Parasuraman, A., 1990, "Five Imperatives for Improving Service Quality" Sloan Management Review, Summer.
- Bowman, C, 1991, Personal Communication, Cranfield
- Collier, D.A., 1987, "The Customer Service and Quality Challenge", Service Industries Journal, Vol 7, No 1
- Gronroos, C, 1984, "A Service Quality Model and Its Marketing Implications", European Journal of Marketing, Vol 18, No.4
- Hart, C.W.L., Heskett, J.L., Sasser, W.E., 1990, "The Profitable Art of Service Recovery", Harvard Business Review, July-August
- Heskett, J.L., Sasser, W.E., Hart, C.W.L., 1990, "Service Breakthroughs", Free Press, New York, p 135
- Lockyer, K, Muhlemann, A., Oakland, J., 1988, "Production and Operations Management", 5th Ed., Pitman.
- Lovelock, C.H., 1984, "Services Marketing", Prentice Hall Inc
- Rhyne, D., Service Industries Journal 1988 "The Impact of Demand Management on Service System Performances", Vol 8, No.4
- Sasser, W.E., 1976, "Match Supply and Demand in Service Industries", Harvard business Review, November-December
- Slack, N, 1987, "Flexibility and The Manufacturing Function - Ten Emperical Observations", Proc. Operations Manager Conference UK, 1987.