Understanding the Queuing Theory for Improved Service Delivery: An Entrepreneurial-Mindset Approach

Tolutope O. Fakokunde (PhD)¹, Adeniyi Mudashiru Mustapha (PhD)² and M. Aremu (PhD)³

Abstract

Queuing theory plays a huge role in solving and preventing operational bottlenecks and service failures in the organization. But, observations show that Queuing analysis may become extremely complex and cumbersome. Therefore, most entrepreneurs, having realized its usefulness in managing their much earned successes achieved in relation to customer’s patronage and service delivery now desire its simplified applications in normal day to day running of operations. The methodology adopted in this paper, therefore is to describe queuing theory and its associated terminologies in relation to service delivery. In view of this, the paper presented a simplified exposition of queuing theory and management of waiting lines as it affects entrepreneurial drive for more business growth and opportunities with its attendant implications to customers’ service delivery and satisfaction. The paper concluded that if appropriately delivered and applied, queuing theory goes a long way in achieving and maintaining customer satisfaction. It recommended that entrepreneurs should seek the opportunity of gaining better understanding and application of queuing theory to practically reduce or eliminate boredom, irritation, breakdown and frustration to customers without much ado.

Keywords: queues, customer, waiting line, entrepreneurs, service.

DOI: https://dx.doi.org/10.4314/ejbe.v6i1.2

¹Department of Entrepreneurship Management Technology, Federal University of Technology, Akure, Nigeria, E-mail: viptopefaks@gmail.com., tofakokunde@futa.edu.ng.
²Department of Business Administration, University of Ilorin, Nigeria
³Department of Business Administration, University of Ilorin, Nigeria
Entrepreneurial - Mindset Approach to Queuing Theory

1. Introduction

The world of Entrepreneurship is a moving target of opportunities, growths and changes (Keogh, 2002; Ferrell, Hirt and Ferrell, 2008). While entrepreneurs are considered as innovators and change agents (Schumpeter, 1942; Nwachukwu, 1990; Drucker, 2010), innovation, according to Weihrich, Cannice and Koontz (2011) is the goal-oriented change to utilize the enterprise’s potential and improve current situations. As the prime source of resources and authority in an economic organisation (Arowomole, 2001), entrepreneurs are purposeful, right thinking, proactive and unassuming in pursuits of identified goals (Awe, 2006) and always willing to respond and adapt to changes by initiating appropriate actions to ensure success (Walske and Zacharakis, 2009) and new order (Arowomole, 2001). According to Awe (2006), the entrepreneur seeks continuous change and undergoes a process of improvement in his organisation and in the lives of the people around him. The process according to Neck and Greene (2011) involves risk taking, identifying an opportunity, understanding resource requirements, acquiring resources, planning and implementing. In ventures creation activities, the entrepreneur perceives needs and conceives new opportunities for expansion (Sagagi, Anyanwu, Aliu and Abimbola, 2014). Thus, the entrepreneur has an instrumental role related to managing creativity and innovation (Stokes, Wilson and Mador, 2010), which may also be defined in terms of innovative outcome to the society (Schindehutte and Morris, 2009). In this light, Greene (2002) sees entrepreneurs as increasingly being pivotal to the fortunes of the British economy. But most often, entrepreneurial profit-oriented search for continuous growth and new opportunities in the market place are hindered by operational bottlenecks, one of which is the menace of queue situations also known as waiting lines. This is not unexpected, as efforts of the entrepreneur to identify and create new methods of working and doing business may cause imbalance between operational capacity and demand at any point in time.

Waiting in lines, according to Stevenson (2009), neither add to customers’ enjoyment nor generate additional revenues for the enterprise, hence, the need to reduce waiting in lines. Also, waiting lines as observed by Awodun and Jongbo (2000) is a common phenomenon which occurs whenever customers arrive randomly for services. According to Awodun and Jongbo (2000), the more society becomes interdependent psychologically, economically and technically, the more individuals encounter waiting lines, or queues, in their daily lives. While, queues may also abound in all sorts of service and production systems, Stevenson (2009) considers it as non - value added occurrence with serious implications to the organization. Thus, queues, if neglected and carelessly handled, may develop to interfere with flourishing business processes and
Entrepreneurial - Mindset Approach to Queuing Theory

existing order, especially when entrepreneurial activities are already rising to its peak (Aremu, 2005). Aremu (2005) asserted that at a point it could make customers to be dejected and dissatisfied and eventually leads to rejection of the enterprise and collapse of the business venture involved. However, Adedayo, Ojo and Obamiro (2010) asserted that the good news is that as long as customers’ patronage increases and waiting lines or queues begin to form, operations and activities involved can still be managed effectively to maintain customers satisfaction with the aids of what is known and consider in management science as Queuing theory.

In the view of Awodun and Jongbo (2000), the entrepreneur, right from inception needs to conveniently and appropriately analyze waiting lines problems with the intention of determining and improving the appropriate level of service. That is, determining the optimal service facilities or the optimal speed of a facility as business growth is achieved or expected. The entrepreneurs, as innovators and change agents, should therefore be wary of such simple unwanted situations that may provoke customers’ complaints and dissatisfaction with services patterns especially as successes started to be recorded in production, patronage and sales. In other words, to manage, solve and prevent crisis situations such as congestion, build-ups, overload, overcrowding, service delays, production bottlenecks, idleness and other similar problems, the entrepreneur, without much ado requires the knowledge and simplified applications of the Queuing theory, which, mathematically, can be complex. This view is supported by (Ferrell, Hirt and Ferrell, 2008) by asserting that the focus on understanding the Queue theory and ability to manage the waiting lines effectively is necessitated by the need for constant improvement of products/services made to ensure continuous patronage and attract new customers.

The objective of the paper therefore is to examine and simplify the concept of queuing theory in relation to its effective applications by entrepreneurs. It therefore explored the basic techniques of queuing theory as an effective tool of customer service delivery in business organisations.

2. Queuing Theory: Definition and Meaning

A queue, according to Sharma (2009), is formed at any place when a customer (human beings or physical entities) that requires service is made to wait due to the fact that the number of customers exceeds the number of service facilities or when service facilities do not work efficiently and take more time than prescribed to serve a customer. Queues, also referred to as waiting in lines are common situations that occur in everyday life (Sharma, 2009). Queues are daily
Entrepreneurial - Mindset Approach to Queuing Theory

encountered in bus stops, banks, supermarkets, saloons, parks, warehouses, airports, clinics, filling stations, traffic lights and so on. In some instances, the customers involved are not people but physical entities or orders to be filled such as trucks waiting to be filled, jobs waiting to be processed or assigned to machines, incoming calls waiting to be picked, vehicles waiting at car wash etc. The need for all these activities to be operated smoothly and efficiently, according to Adam Jr. and Ebert (2001) arose the use of Queuing theory, which although limited in complex situations is better than using human intuition. Queuing theory is a mathematical approach to the analysis of waiting lines with varied applications in service operations. It is a branch of management science that enables the analyst to describe and understand the behavior of a system as reflected in its operating characteristics (Awodun and Jongbo, 2000). It can be applied to situations where it is now possible to predict accurately the time/rate of arrivals of customers and time/rate of service. Queuing theory utilizes mathematical models and performance measures to assess and hopefully improve the flow of customers through a queuing system (Prabhu, 1997 and Gorney, 1981). Queuing theory has been used in the past to assess such things as staff schedules, working environment, productivity customer waiting time and customer waiting environment (Ronald and James, 2001).

2.1. Reasons for Queues
A queue is typically formed whenever customers arrived and the facility is busy (Awodun and Jongbo, 2000). A situation when limited service facilities fail to satisfy the demands for service that are made upon them brings bottlenecks and thereafter generate queue or waiting line (Aminu, 2000). A similar opinion was expressed by Michael (2001) when he submitted that in situations where facilities are limited and cannot satisfy the demand made upon them, bottlenecks occur which manifest as queue but customers are not interested in waiting in queues. When service providers are very small, queues will also arise (Aremu, 2005). The views expressed above clearly show that inadequate facility is a major reason for customers waiting, especially in situations when arrivals can be scheduled and service made constant. Similarly, when analyzing various forms of queues, Trueman (1977) and Slack, Chambers and Johnston (2010) stated that, queues are formed when units receiving some types of service cannot be served immediately. The type of units being served, the actual service performed, and the queue itself can take many different forms as shown in the example displayed in Table 1 below:
Entrepreneurial - Mindset Approach to Queuing Theory

Table 1: Typical Examples of Queuing Activities.

<table>
<thead>
<tr>
<th>Arriving units</th>
<th>Servers</th>
<th>Queues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplanes</td>
<td>Runways</td>
<td>Plane waiting to land or take off</td>
</tr>
<tr>
<td>Ships</td>
<td>Repair docks</td>
<td>Ships awaiting repair</td>
</tr>
<tr>
<td>Computer jobs</td>
<td>Computer</td>
<td>Job waiting processing</td>
</tr>
<tr>
<td>Budgetary request</td>
<td>Budget</td>
<td>Unfounded requests</td>
</tr>
<tr>
<td>Criminal cases</td>
<td>Criminal Courts</td>
<td>Cases awaiting trial</td>
</tr>
<tr>
<td>Library patrons</td>
<td>Books</td>
<td>Waiting list</td>
</tr>
<tr>
<td>Order of Merchandise</td>
<td>Inventory</td>
<td>Book orders</td>
</tr>
</tbody>
</table>

Sources: Adapted from Trueman (1977) and Slack et al (2010).

The waiting lines developed because the service to customer may not be rendered immediately as the customer reaches the service facility (Vohra, 2007). According to Slack, Chambers and Johnston (2010), only rarely will the arrival of customers match the service capacity to cope with them. Thus, lack of adequate service facility would cause waiting lines of customers to be formed. The only way that the service demand can be met with ease, according to Vohra (2007) is to increase the service capacity (and raising the efficiency of the existing capacity if possible) to a higher level. This may have informed Trueman (1977) submission that queuing problem arises primarily because of economic consideration, which indeed, is a rare situation in which the cost of service is so low that enough service facilities can be provided so that no arrival has to wait. The only situation of this nature that comes to mind is that of self service.

Queues may also occur even if service system can provide service at faster rate than customer's arrival rate, but the arrival and service processes are random (Ashley, 2000). Stevenson (2009) added that the high degree of variability usually exhibited by service and arrival patterns may cause facility system to be temporarily overloaded, as different from situations in which queue is formed when arrivals are scheduled and service constant. In this case, waiting lines can still form at a point when both arrivals and service can no longer be scheduled and held constant respectively and the facility thus becomes inadequate. This may probably explain why waiting lines or congestions are usually experienced at the end of every month in most bank premises in Nigeria due to the possibility of huge volume of transactions during the period. These delays used to occur because it is not possible to pre-determine the arrival behaviour of customers during these periods.
James and Benson, (1988) in their own contributions, posited that the formation of waiting lines (queues) is a phenomenon that occurs whenever the demand for a service exceeds its supply. They expressed that if long queues develop, it may be an indication that not enough service is being provided. If no queues develop, it may be an indication that too much service is being provided. Either situation can prove costly to the service provider.

It now becomes clearer that queues or waiting lines are inevitable in organizations and are basically formed or may develop due to the following reasons:

- Units arriving at a facility to receive a particular type of service but cannot be serviced immediately. In this case, Traffic Intensity should be used. This is the ratio of the average arrival rate to the average service rate. It is commonly assumed that the traffic intensity of a queue is less than one.
- When the service time is less than the arrival time. Wild (1980) maintained that if the service rate ($\mu$) is less than the arrival rate ($\lambda$) the queue elongates.
- Randomness of arrival and service rates.
- When the demand for a service facility exceeds the supply.

In addition to the above, queues may also develop due to the occurrence of some of the following circumstances:

(i) Temporary or permanent breakdown or withdrawal of service facilities.
(ii) Queuing units preference for a particular service facility leading to partial queues for the preferred facility.
(iii) Work attitude of (human) servers. Hannagan (1995) with regards to workers attitude submits that all changes involve people and their working patterns and that organizations can only change at the speed at which people in them are willing and able to change.

2.2. The Structure of the Queuing System

The entire structure of queuing systems is usually presented as measuring efficiency and effectiveness of a system in terms of the number of customers in the system and in queues, the average time customers spent in the system and the probability of the system being busy or idle (Stevenson, 2009). More often, this involves the use of computational figures to arrive at well-calculated deductions without recourse to the cost of operating the system. In this manner, organizations where customers only spent few minutes in the system and in the queues are erroneously adjudged to be operating smoothly and efficiently. Hence, it is now important for entrepreneurs, managers and other system
Entrepreneurial - Mindset Approach to Queuing Theory

operators to weigh the cost of providing a given level of service capacity against the potential cost of having customers to wait for service (Fakokunde, 2002; Aremu, 2005; Stevenson, 2009). In this direction, more relevant and appropriate is the structure of the Queuing system presented by Prabhu (1997) which comprises the characteristics of the following elements: the arrival, the queue discipline, the service mechanism, and the cost structure. This simply involves the inclusion of the cost elements into the components of a typical queue system showing in Figure 1.

Figure 1: Components of a Queuing System.

Source: Sharma (2009).

According to Prabhu (1997), the arrival process is the manner in which customers enters the system for service delivery. It also refers to the way in which customers show up at the service facility (Awodun and Jongbo, 2000). The number of potential customers could be finite or infinite while arrival pattern could be either random or in an organized pattern (Daellenbach and George, 1978). The analysis of queues being considered is that of infinite calling population since in most cases, the arrival rate may not be determined and constant. Thus, the arrival rate could basically be assumed to be random. Usually an infinite calling population can be assumed if the rate of arrival of future customers is not affected by the number of customers already on the queue. When arrivals are random, it is required that the probability distribution describing specifically inter-arrival times must be known. It is interesting for entrepreneurs to note that customers exhibit different attitudes in the waiting area. While a customer decides to wait patiently regardless of the number in queue until served, another customer may be impatient and seeks service elsewhere. The varying attitudes as identified by (Sharma 2009) are Reneging, Balk, or Jockeying. Arriving customers are expected to approach the service system, and wait for service regardless of the number in queue. But, according to Davis and Heinete (1994), a balking customer dissatisfied with the queue, refuses to join and may or may not return later, while a reneging customer, although joined the queue, wait for some time, becomes tired and impatient, then leaves. Customers who move from one queue to another hoping to receive
Entrepreneurial - Mindset Approach to Queuing Theory

service more quickly are said to be jockeying. Also, when arrivals collaborate to reduce overall waiting time, the behaviour is termed collaborate.

The queue discipline refers to which, and what units in the calling population receives what service, how and when (Daellenbach and George, 1978). The queue discipline describes the rules and policies guiding the operation of the waiting line or queue. It refers to the order or manner in which jobs are processed or customers in the queue are served. It may be on a first-come-first-served, random or subject to service priorities on the basis of some attributes such as urgencies and emergencies. The first-come-first served queue disciplines are very much common in most contemporary service systems e.g. banks, supermarkets, restaurants and hospitals.

In addition, priority discipline can be regrouped into two, i.e. pre-emptive and non-pre-emptive. Pre-emptive discipline permits and allows units to interrupt/break the continuity of units already receiving service. On the contrary, non-pre-emptive queue discipline arranges the queue so that the unit with the greatest priority rating gets served before others. Also, when units requiring service receive programs being run on time-shared terminals, we have a round-robin queue discipline. There may also be attempts by some customers to flout the rules guiding the service delivery procedures (Vohra, 2007). In this manner, entrepreneurs are to ensure proper monitoring of the waiting area to prevent acts such as bribing and cheating by some customers for queuing position.

The service mechanism describes how the customer is served. The service facility may consist of one or several situations or channels. They may operate either in parallel, in which case an arrival has to go through one channel only before being discharged from the system, or they may operate in series, in which case an arrival has to go through several channels in sequence before being discharged. Decision on the structure and nature of queues would be incomplete and remain an arm chair academic exercise without considering the probability distribution describing the service times (Griffin, 1978). The service times at each channel may be constant or random with a known service time distribution.

According to Vohra (2007), there are two aspects of a service system which the entrepreneur must consider:

(i) the structure of the service system, and
(ii) the speed of service.

The structure of the service system means how the service facilities exist. There may be a single service facility, a multiple, parallel facilities with single queue, multiple parallel facilities with multiple queues and services facilities in a series. In a queuing system, the speed with which service is provided can be expressed in either of two ways - as service rate and service time. The service rate describes
the number of customers served during a particular time, while the service time indicates the amount of time needed to service customers. Service rates and service times are reciprocals of each other and either of them is sufficient to indicate the capacity of the facility. In determining a particular capacity level of operation, it is incumbent on the entrepreneur to identify the cost of either increase the service rate or reduce the service time. There is also sojourn time. This is waiting time plus the service time.

The cost structure specifies the payment made by the customer and the various operating costs of the system. The goal of waiting line management is essentially to minimize total cost involved in operating the system. The entrepreneur while trying to remove idle time needs to balance the cost of offering an acceptable level of service capacity with the cost of customers waiting, and even refusing to wait for service due to delay. According to Sharma (2009), the more the capacity increases, the less the waiting customers, time and of course, cost. It is important for the entrepreneur to identify a level of service capacity at which cost will be minimized. The optimum service capacity level is identified by Sharma (2009) as the one that minimizes the sum of the two costs. Thus,

\[ TC = \text{Customer waiting cost} + \text{Capacity cost} \]

From the basic and fundamental relationship stated above, one can represent the analysis graphically thus:

Figure 2: Relationship between Level of Service and Cost of Waiting Line.

Source: Adapted from Griffin (1978)

It is essential that cost of waiting and cost of service should be such as to minimize total expected cost, and while experience and mathematical formulae
Entrepreneurial - Mindset Approach to Queuing Theory

may be used for simple waiting line situations, complex situations can be solved by simulation methods (Olaitan, Bojerenu and Onyebuchi, 2015).

The optimum service facility that minimizes costs, according to Griffin (1978) is also given as:

\[
\mu = \lambda + \frac{Cw\lambda}{Cs}
\]

(1)

where
\[
\mu = \text{Average service rate}
\]
\[
\lambda = \text{Average arrival rate}
\]
\[
Cw = \text{Cost of waiting in the queue.}
\]
\[
Cs = \text{Cost of service per unit of time}
\]

The cost structure analysis involves the use of both explicit and implicit costs elements (Aremu, 2005). Queuing models that restrict analysis of waiting line situations to the use of explicit cost are not comprehensive enough and less satisfactory. Implicit cost elements such as frustrations, boredom, irritation, man-hour loss and associated hazard arising from waiting in line are therefore essential elements to be taken into consideration (Olaitan, Bojerenu and Onyebuchi, 2015).

3. Managing the Waiting Line and Customer Satisfaction

Stevenson (2009) mentioned that managers have a number of very good reasons to be concerned with waiting lines. Chief among these reasons are:

i. the cost implication of providing waiting space.
ii. the loss of business as a result of customers reneging and balking.
iii. the loss of goodwill.
iv. the disturbance to or disruption of other business activities and customers.
v. a possible reduction in customers satisfaction.

The dilemma of managing queues as observed by Slack, Chambers and Johnston (2010) occurs as a result of the probabilistic arrival and processing times, which in reality are difficult to predict accurately. Managing the customer's perception of the queuing experience can then be the vital element needed for satisfactory service interaction to be achieved. Perception is considered by Moore, Petty, Palich and Longenecker (2010) as one of the four psychological factors having greatest influence on customers’ behavior. Others are needs, motivations and
Entrepreneurial - Mindset Approach to Queuing Theory

attitudes. It encompasses of those individual processes that ultimately give meaning to the stimuli consumer encounter. Slack, Jones and Dent (1994) and Chambers and Johnston (2010) suggested that customers judge the service they receive on how they perceive the time spent queuing, and that there is a point where a lengthy wait begins to affect the customer's perception of quality. The measurement of customer satisfaction as it relates to waiting time is therefore highly qualitative and subjective, and the relationship is naturally inverse (i.e. in general, as waiting time decreases, satisfaction increases). Achieving customer satisfaction is therefore important for most marketers and consumer researchers (Fournier and Mick, 1999) and entrepreneurs alike.

To measure therefore, the effectiveness of any queuing system, two categories of statistics are classified by Gorden and Pressman (1975); the users-oriented statistics and system-oriented statistics. A user-oriented statistics measures performance related to what the users experience while the system-oriented statistics measures the efficiency of the system.

According to Gorden and Pressman (1975) the user-oriented statistics can be used to measure the:
(a) waiting time. That is, the time spent waiting in the queue for the service,
(b) time spent in the system, and
(c) likelihood (probability) of a user leaving without being served, if there is balking, reneging, or a limited queue.

From the user’s point of view, a system that provides, on the average smaller waiting time, less time spent in the system, or low likelihood of waiting is providing better service. But, it is interesting to note that the presence of a queue may not after all attract bad recommendation especially in the use and patronage of public goods and utilities. It may depend on the perception of customers with regards to service delivery. At times, it is commonly understood that regular queues in a public hospital or even a private clinic signifies that patients are served on time, the workers are of good attitude and the medical attention being given is good. Also, regular queue at a cinema theatre implies that the shows are worth seeing. Similarly, queues in banking hall and petrol stations might be desirable at times, and attest to repeat purchase which can be attributed to the presence of good service delivery. Most modern technologies in factories and the work place now operate on the basis of queuing principles in order to reduce or totally eliminate idle time which may ensue if there are no queues.

These present day realities have therefore made everybody to be familiar with queue situations, thus the notion ‘to find satisfaction, look for a queue to join.’ This notion may be attributable to the fact that service improvement efforts may initially attract queues and additional lines preferred by customers waiting with
the expectations of more satisfaction. In this direction, a highly satisfied customer will very likely provide repeat business and spread the positive experience by word of mouth (advertising) while on the other hand, a dissatisfied customer will most likely not provide repeat business and will be more than willing to share his or her bad experience with whoever will listen. Nevertheless, in any case, the entrepreneur, who is interested in operating the system at a desirable optimum level, must endeavour to set up a system to counteract any bad experiences customers may experience (Moore, Petty, Palich and Longenecker, 2010), avoid endless and explosive queue situations as well as attempt to manage customers’ perceptions and expectations since the experience of each user with respect to service delivery will not be the same.

4. Conclusion

No doubt, Queuing theory as an analytical tool in management science has contributed all along to the applications of waiting line knowledge in organizations with the general view that the queue relationship with customer satisfaction is inverse in nature. In doing so, an organization where customers spent few minutes in the system and in the queues are adjudged to be most efficient and effective. Rather than dwell more on this conventional model of analysing the queuing system, the entrepreneur prefers to be more customer-focused and profit-driven by taking all the full cost element into consideration in managing the waiting lines. It is therefore pertinent for the entrepreneur to seek and desire a clear understanding and application of queuing theory to practically create efforts that add value to the system while at the same time reduce or eliminate boredom, irritation, frustration, breakdowns and other crisis situations, known fully well that a satisfied customer will very likely provide repeat business and spread the positive experience and a dissatisfied customer will act otherwise.

The paper has therefore shown that managing the customer’s perception of the queuing experience appropriately, can therefore be the vital element which effective customer service delivery and satisfaction demands. Thus, Queuing theory and management of waiting lines should be of considerable interest to entrepreneurs and all business owners alike in rational analysis of day to day operations and decision making.
Entrepreneurial - Mindset Approach to Queuing Theory

References


Greene, F. J. (2002). An Investigation into Enterprise Support for Younger
Entrepreneurial - Mindset Approach to Queuing Theory

Entrepreneurial - Mindset Approach to Queuing Theory


