Chapter 5

Service Design

Russell and Taylor
Operations and Supply Chain Management, 8th Edition
Lecture Outline

- **Service Economy** – Slide 4
- **Characteristics of Services** – Slide 8
- **Service Design Process** – Slide 11
- **Tools for Service Design** – Slide 18
- **Waiting Line Analysis for Service Improvement** – Slide 21
Learning Objectives

• Evaluate the impact of services on jobs and the economy
• Appreciate and articulate the differences between products and services
• Utilize tools for envisioning and designing quality services
• Map out service processes and suggest process improvements
• Model waiting lines and evaluate their performance for service improvement
Service Economy

International Employment by Industry Sector

© 2014 John Wiley & Sons, Inc. - Russell and Taylor 8e
U.S. Economy – Percent of GDP by Industry Sector
Characteristics of Services

- **Services**
  - acts, deeds, or performances
- **Goods**
  - tangible objects
- **Facilitating services**
  - accompany almost all purchases of goods
- **Facilitating goods**
  - accompany almost all service purchases
Continuum From Goods to Services

<table>
<thead>
<tr>
<th>Goods</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Automobile purchase</td>
<td>Rental car</td>
</tr>
<tr>
<td>Take-out food</td>
<td>Auto repair</td>
</tr>
<tr>
<td>Restaurant meal</td>
<td>Cell phone</td>
</tr>
<tr>
<td>Hospital</td>
<td>Cell phone</td>
</tr>
<tr>
<td>Massage</td>
<td>Tax preparation</td>
</tr>
<tr>
<td>Lawn cutting</td>
<td>Lawn cutting</td>
</tr>
</tbody>
</table>
Characteristics of Services

- Services are intangible
- Service output is variable
- Services have higher customer contact
- Services are perishable
- The service and the service delivery are inseparable
- Services tend to be decentralized and geographically dispersed
- Services are consumed more often than products
- Services can be easily emulated
Service Design Process
Service Design Process

• Service concept
  • purpose of a service; it defines target market and customer experience

• Service package
  • mixture of physical items, sensual benefits, and psychological benefits

• Service specifications
  • performance specifications
  • design specifications
  • delivery specifications
## High vs. Low Contact Services

<table>
<thead>
<tr>
<th>Design Decision</th>
<th>High-Contact Service</th>
<th>Low-Contact Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility location</td>
<td>Convenient to customer</td>
<td>Near labor or transportation source</td>
</tr>
<tr>
<td>Facility layout</td>
<td>Must look presentable, accommodate customer needs, and facilitate interaction with customer</td>
<td>Designed for efficiency</td>
</tr>
</tbody>
</table>
### High vs. Low Contact Services

<table>
<thead>
<tr>
<th>Design Decision</th>
<th>High-Contact Service</th>
<th>Low-Contact Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality control</td>
<td>More variable since customer is involved in process; customer expectations and perceptions of quality may differ; customer present when defects occur</td>
<td>Measured against established standards; testing and rework possible to correct defects</td>
</tr>
<tr>
<td>Capacity</td>
<td>Excess capacity required to handle peaks in demand</td>
<td>Planned for average demand</td>
</tr>
</tbody>
</table>
High vs. Low Contact Services

<table>
<thead>
<tr>
<th>Design Decision</th>
<th>High-Contact Service</th>
<th>Low-Contact Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Worker skills</td>
<td>▪ Must be able to interact well with customers and use judgment in decision making</td>
<td>▪ Technical skills</td>
</tr>
<tr>
<td>▪ Scheduling</td>
<td>▪ Must accommodate customer schedule</td>
<td>▪ Customer concerned only with completion date</td>
</tr>
</tbody>
</table>
# High vs. Low Contact Services

<table>
<thead>
<tr>
<th>Design Decision</th>
<th>High-Contact Service</th>
<th>Low-Contact Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service process</td>
<td>Mostly front-room activities; service may change during delivery in response to customer</td>
<td>Mostly back-room activities; planned and executed with minimal interference</td>
</tr>
<tr>
<td>Service package</td>
<td>Varies with customer; includes environment as well as actual service</td>
<td>Fixed, less extensive</td>
</tr>
</tbody>
</table>
Tools for Service Design

• Service blueprinting
  • line of influence
  • line of interaction
  • line of visibility
  • line of support
• Front-office/Back-office activities

• Servicescapes
  • space and function
  • ambient conditions
  • signs, symbols, and artifacts
• Quantitative techniques
Service Blueprinting
Service Blueprinting

Customer passes “Today’s Special” sign

Customer places order

Barista readies order

Barista goes to stockroom, gets last bag of cups

Barista tells manager out of cups

Barista asks for help upfront

Tour bus stops

Customer pays

Customers pour in

Line of Influence

Line of Interaction

Line of Visibility

Line of Support
Elements of Waiting Line Analysis

- Operating characteristics
  - average values for characteristics that describe performance of waiting line system
- Queue
  - a single waiting line
- Waiting line system
  - consists of arrivals, servers, and waiting line structure
- Calling population
  - source of customers; infinite or finite
Elements of Waiting Line Analysis

• Arrival rate ($\lambda$)
  • frequency at which customers arrive at a waiting line according to a probability distribution, usually Poisson

• Service rate ($\mu$)
  • time required to serve a customer, usually described by negative exponential distribution

• Service rate must be higher than arrival rate ($\lambda < \mu$)

• Queue discipline
  • order in which customers are served

• Infinite queue
  • can be of any length; length of a finite queue is limited
Elements of Waiting Line Analysis

- **Channels**
  - number of parallel servers for servicing customers

- **Phases**
  - number of servers in sequence a customer must go through
Operating Characteristics

- Operating characteristics are assumed to approach a steady state

<table>
<thead>
<tr>
<th>Notation</th>
<th>Operating Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Average number of customers in the system (waiting and being served)</td>
</tr>
<tr>
<td>$L_q$</td>
<td>Average number of customers in the waiting line</td>
</tr>
<tr>
<td>$W$</td>
<td>Average time a customer spends in the system (waiting and being served)</td>
</tr>
<tr>
<td>$W_q$</td>
<td>Average time a customer spends waiting in line</td>
</tr>
<tr>
<td>$P_0$</td>
<td>Probability of no (i.e., zero) customers in the system</td>
</tr>
<tr>
<td>$P_n$</td>
<td>Probability of $n$ customers in the system</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Utilization rate; the proportion of time the system is in use</td>
</tr>
</tbody>
</table>
Traditional Cost Relationships

- As service improves, cost increases

![Graph showing cost relationships](image-url)
Psychology of Waiting

- Waiting rooms
  - magazines and newspapers
  - televisions
- Bank of America
  - mirrors
- Supermarkets
  - magazines
  - “impulse purchases”
Psychology of Waiting

• Preferential treatment
  • Grocery stores: express lanes for customers with few purchases
  • Airlines/Car rental agencies: special cards available to frequent-users or for an additional fee
  • Phone retailers: route calls to more or less experienced salespeople based on customer’s sales history

• Critical service providers
  • services of police department, fire department, etc.
  • waiting is unacceptable; cost is not important
Waiting Line Models

- **Single-server model**
  - simplest, most basic waiting line structure
- Frequent variations (all with Poisson arrival rate)
  - exponential service times
  - general (unknown) distribution of service times
  - constant service times
  - exponential service times with finite queue
  - exponential service times with finite calling population
Basic Single-Server Model

• Assumptions
  • Poisson arrival rate
  • exponential service times
  • first-come, first-served queue discipline
  • infinite queue length
  • infinite calling population

• Computations
  • $\lambda = \text{mean arrival rate}$
  • $\mu = \text{mean service rate}$
  • $n = \text{number of customers in line}$
Basic Single-Server Model

- probability that no customers are in queuing system
  \[ P_0 = \left( 1 - \frac{\lambda}{\mu} \right) \]

- probability of \( n \) customers in queuing system
  \[ P_n = \left( \frac{\lambda}{\mu} \right)^n \cdot P_0 = \left( \frac{\lambda}{\mu} \right)^n \left( 1 - \frac{\lambda}{\mu} \right) \]

- average number of customers in queuing system
  \[ L = \frac{\lambda}{\mu - \lambda} \]

- average number of customers in waiting line
  \[ L_q = \frac{\lambda^2}{\mu (\mu - \lambda)} \]
Basic Single-Server Model

- average time customer spends in queuing system
  
  \[ W = \frac{1}{\mu - \lambda} = \frac{L}{\lambda} \]

- average time customer spends waiting in line
  
  \[ W_q = \frac{\lambda}{\mu (\mu - \lambda)} \]

- probability that server is busy and a customer has to wait (utilization factor)
  
  \[ \rho = \frac{\lambda}{\mu} \]

- probability that server is idle and customer can be served
  
  \[ I = 1 - \rho = 1 - \frac{\lambda}{\mu} = P_0 \]
Basic Single-Server Model Example

\[ \lambda = 24 \]
\[ \mu = 30 \]

\[ P_0 = \]

\[ L = \]

\[ L_q = \]
Basic Single-Server Model Example

\[ \lambda = 24 \]
\[ \mu = 30 \]

\[ P_0 = \left(1 - \frac{\lambda}{\mu}\right) = \left(1 - \frac{24}{30}\right) \]
\[ = 0.20 \text{ probability of no customers in the system} \]

\[ L = \frac{\lambda}{\mu - \lambda} = \frac{24}{30 - 24} \]
\[ = 4 \text{ customers on the average in the queuing system} \]

\[ L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{(24)^2}{30(30 - 24)} \]
\[ = 3.2 \text{ customers on the average in the waiting line} \]
Basic Single-Server Model Example

\[ W = \]
\[ W_q = \]
\[ \rho = \]
\[ I = \]
Basic Single-Server Model Example

\[ W = \frac{1}{\mu - \lambda} = \frac{1}{30 - 24} \]

= 0.167 hour (10 minutes) average time in the system per customer

\[ W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{24}{30(30 - 24)} \]

= 0.133 hour (8 minutes) average time in the waiting line per customer

\[ \rho = \frac{\lambda}{\mu} = \frac{24}{30} \]

= 0.80 probability that the server will be busy and the customer must wait

\[ I = 1 - \rho = 1 - 0.80 \]

= 0.20 probability that the server will be idle and a customer can be served
Service Improvement Analysis

• Waiting time (8 min.) is too long
  • hire assistant for cashier?
    • increased service rate
  • hire another cashier?
    • reduced arrival rate
• Is improved service worth the cost?
### A Single-Server Model

#### Example 5.1

**Input:**
- Arrival rate = 24 per hour
- Service rate = 30 per hour

**Output:**
- Average number in the system (L) = 4.00
- Average number in the queue (Lq) = 3.20
- Average time in the system (W) = 10.00 minutes
- Average time in the queue (Wq) = 8.00 minutes
- Utilization factor (U) = 0.800
- P(0) = 0.200

- Number in the system, n = 5
- P(n) = 0.066

**Calculations:**

- Lq, number in queue: $Lq = \frac{L}{W - \lambda}$
- P(n): $P(n) = \frac{P_{0}^{n}}{n!}$

**Notes:**
- Input the arrival rate and service rate. If given time between arrivals or service time, divide into 60 minutes to convert to rates.
- Input n to find the probability of n customers in the system.
Advanced Single-Server Models

• Constant service times
  • occur most often when automated equipment or machinery performs service

• Finite queue lengths
  • occur when there is a physical limitation to length of waiting line

• Finite calling population
  • number of “customers” that can arrive is limited
Advanced Single-Server Models

(a) Single-Server, Constant Time

(b) Single-Server, Finite Queue

(c) Single-Server, Finite Calling Population
## Advanced Single-Server Model

### A Single-Server Model with Finite Queue

**Exhibit 5.2**

**Input:**
- Arrival rate = 10 per hour
- Service rate = 7 per hour
- M = 5

**Output:**
- \( P_0 = 0.06 \)
- \( P_n = 0.34 \)
- Average number in the system (L) = 3.47
- Average number in the queue (Lq) = 2.52
- Average time in the system (W) = 31.52 minutes
- Average time in the queue (Wq) = 22.94 minutes

*Input the arrival rate, service rate, and the maximum number in the system.*

Probability of zero customers
Basic Multiple-Server Model

• Single waiting line and service facility with several independent servers in parallel
• Same assumptions as single-server model
• $s \mu > \lambda$
  • $s = \text{number of servers}$
  • servers must be able to serve customers faster than they arrive
Basic Multiple-Server Model

• probability that there are no customers in system

\[ P_0 = \frac{1}{\sum_{n=0}^{n=s-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n + \frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^s \left( \frac{s\mu}{s\mu - \lambda} \right)} \]

• probability of \( n \) customers in system

\[ P_n = \begin{cases} \frac{1}{s!s^{n-s}} \left( \frac{\lambda}{\mu} \right)^{n-s} P_0, & \text{for } n > s \\ \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n P_0, & \text{for } n \leq s \end{cases} \]
Basic Multiple-Server Model

- probability that customer must wait

\[ P_w = \frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^s \frac{s \mu}{s \mu - \lambda} P_0 \]

\[ L_q = L - \frac{\lambda}{\mu} \]

\[ L = \frac{\lambda \mu (\lambda/\mu)^s}{(s - 1)! (s \mu - \lambda)^2} P_0 + \frac{\lambda}{\mu} \]

\[ W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda} \]

\[ W = \frac{L}{\lambda} \]

\[ \rho = \frac{\lambda}{s \mu} \]
Basic Multiple-Server Model Example

Three-server system

\( \lambda = 10/\text{hr} \)

\( \mu = 4/\text{hr} \)

\( S = 3 \)

\( s\mu = 3 \times 4 = 12 \)

\[
P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n + \frac{1}{s!} \frac{\lambda}{\mu} \left( \frac{s\mu}{s\mu - \lambda} \right)}
\]

\[
= \frac{1}{\left[ \frac{0}{0!} \left( \frac{10}{4} \right)^0 + \frac{1}{1!} \left( \frac{10}{4} \right)^1 + \frac{1}{2!} \left( \frac{10}{4} \right)^2 \right] + \frac{1}{3!} \left( \frac{10}{4} \right)^3} \frac{3(4)}{3(4) - 10}
\]

\[
= 0.045 \text{ probability that no customers are in the health service.}
\]
Basic Multiple-Server Model Example

\[ L = \frac{\lambda \mu (\lambda/\mu)^s}{(s - 1)!(s\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu} \]

\[ = \frac{(10)(4)(10/4)^3}{(3 - 1)!(3(4) - 10)^2} \times (0.045) + \frac{10}{4} \]

\[ = 6 \text{ students in the health service} \]

\[ W = \frac{L}{\lambda} \]

\[ = \frac{6}{10} \]

\[ = 0.60 \text{ hour or } 36 \text{ minutes in the health service} \]

\[ L_q = L - \frac{\lambda}{\mu} \]

\[ = 6 - \frac{10}{4} \]

\[ = 3.5 \text{ students waiting to be served} \]

\[ W_q = \frac{L_q}{\lambda} \]

\[ = \frac{3.5}{10} \]

\[ = 0.35 \text{ hour or } 21 \text{ minutes waiting in line} \]
Basic Multiple-Server Model Example

\[ P_w = \frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^s \frac{s\mu}{s\mu - \lambda} P_0 \]

\[ = \frac{1}{3!} \left( \frac{10}{4} \right)^3 \frac{3(4)}{3(4) - (10)} (0.045) \]

\[ = 0.703 \text{ probability that a student must wait for service} \]
\[ \text{(i.e., that there are three or more students in the system)} \]
Basic Multiple-Server Model Example

- To cut waiting time, add another service rep
- Four-server System

\[ P_0 = 0.073 \text{ probability that no students are in the health service} \]
\[ L = 3.0 \text{ students in the health service} \]
\[ W = 0.30 \text{ hour, or 18 minutes, in the health service} \]
\[ L_q = 0.5 \text{ students waiting to be served} \]
\[ W_q = 0.05 \text{ hour, or 3 minutes, waiting in line} \]
\[ P_w = 0.31 \text{ probability that a student must wait for service} \]
Multiple-Server Waiting Line in Excel

A Multiple-Server Waiting Line System

**Input:**
- Arrival rate = 10 per hour
- Service rate = 4 per hour
- No. of servers, s = 3

**Output:**
- \( P_o = 0.045 \)
- \( P_w = 0.702 \)
- Average number in the system (\( L \)) = 6.01
- Average number in the queue (\( L_q \)) = 3.51
- Average time in the system (\( W \)) = 36.07 minutes
- Average time in the queue (\( W_q \)) = 21.07 minutes

Input the arrival rate, service rate, and number of servers.
Multiple-Server Waiting Line in Excel

### Multiple-Server Model

\[ P_0 = \frac{1}{\left[ \sum_{n=0}^{\pi-s-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n \right] + \frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^s \frac{s\mu}{s\mu - \lambda}} \]

\[ P_n = \begin{cases} 
\frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^n P_0, & \text{for } n > s \\
\frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n P_0, & \text{for } n \leq s 
\end{cases} \]

\[ P_w = \frac{1}{s!} \left( \frac{\lambda}{\mu} \right)^s \frac{s\mu}{s\mu - \lambda} P_0 \]

\[ L = \frac{\lambda \mu (\lambda / \mu)^s}{(s-1)! (s\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu} \]

\[ W = \frac{L}{\lambda} \]

\[ L_q = L - \frac{\lambda}{\mu} \]

<table>
<thead>
<tr>
<th>( n )</th>
<th>Summation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>3.5000</td>
</tr>
<tr>
<td>3</td>
<td>6.6250</td>
</tr>
<tr>
<td>4</td>
<td>9.2292</td>
</tr>
<tr>
<td>5</td>
<td>10.8568</td>
</tr>
<tr>
<td>6</td>
<td>11.6706</td>
</tr>
<tr>
<td>7</td>
<td>12.0097</td>
</tr>
<tr>
<td>8</td>
<td>12.1308</td>
</tr>
<tr>
<td>9</td>
<td>12.1686</td>
</tr>
<tr>
<td>10</td>
<td>12.1791</td>
</tr>
<tr>
<td>11</td>
<td>12.1817</td>
</tr>
<tr>
<td>12</td>
<td>12.1823</td>
</tr>
<tr>
<td>13</td>
<td>12.1825</td>
</tr>
<tr>
<td>14</td>
<td>12.1825</td>
</tr>
<tr>
<td>15</td>
<td>12.1825</td>
</tr>
<tr>
<td>16</td>
<td>12.1825</td>
</tr>
<tr>
<td>17</td>
<td>12.1825</td>
</tr>
<tr>
<td>18</td>
<td>12.1825</td>
</tr>
<tr>
<td>19</td>
<td>12.1825</td>
</tr>
<tr>
<td>20</td>
<td>12.1825</td>
</tr>
</tbody>
</table>