INTEGRATION OF SIMULATION AND DEA TO DETERMINE THE MOST EFFICIENT PATIENT APPOINTMENT SCHEDULING MODEL FOR A SPECIFIC CLINIC SETTING

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agricultural and Applied Science

By

Nazanin Aslani

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Industrial and Manufacturing Engineering

June 2011

Fargo, North Dakota

North Dakota State University
Graduate School

Title

INTEGRATION OF SIMULATION AND DEA TO DETERMINE THE MOST EFFICIENT

PATIENT APPOINTMENT SCHEDULING MODEL FOR A SPECIFIC CLINIC SETTING

By

NAZANIN ASLANI

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

ABSTRACT

Aslani, Nazanin, M.S., Department of Industrial and Manufacturing Engineering, College of Engineering and Architecture, North Dakota State University, June 2011. Integration of Simulation and DEA to Determine the Most Efficient Patient Appointment Scheduling Model for a Specific Clinic Setting. Major Professor: Dr. Jun Zhang.

This study develops a method to determine the most efficient scheduling model for a specific clinic setting.

The appointment scheduling system assigns clinics' timeslots to incoming requests. There are three major scheduling models: centralized scheduling model (CSM), decentralized scheduling model (DSM) and hybrid scheduling model (HSM). In order to schedule multiple appointments, CSM involves one scheduler, DSM involves all the schedulers of individual clinics and HSM combines CSM and DSM.

Clinic settings are different in terms of important factors such as randomness of appointment arrival and proportion of multiple appointments.

Scheduling systems operate inefficiently if there is not an appropriate match between scheduling models and clinic settings to provide balance between indicators of efficiency. A procedure is developed to determine the most efficient scheduling model by the integrated contribution of simulation and Data Envelopment Analysis (DEA). A case study serves as a guide to use and as proof for the validity of the developed procedure.

ACKNOWLEDGMENTS

I am heartily thankful to my advisor, Dr. Jun Zhang, for without her knowledge, guidance and encouragement, this study would not have been successful. She guided me from the first semester that I started my master's program and taught me how to do research. Her exceptional talent to develop new research topics and her patience to teach motivated me to work hard and pursue my interest, which is choosing a life based on research and science and continuing my studies for a PhD. Her help is not only limited to the academic area but also my daily life.

I would like to express my gratitude to Dr. Kambiz Farahmand for his advice and guidance during the entire period of my study and supporting me in applying to PhD programs.

I wish to offer my regards and blessings to Dr. Jing Shi, who was supportive of my research and encouraged me to continue my studies for a PhD program, and Dr. Chanchai Tangpong, who supported me and gave me confidence during the completion of the project.

I would also like to give my most heartfelt gratitude to my parents who gave me the confidence to come to the US to study. I express my special gratitude to my love, Fariborz, who has always made my life meaningful. I wish to thank my sister, Saba, who has always been my best friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1. Background

The scheduling of appointments in outpatient clinics is the process of assigning clinics' timeslots to incoming requests (Guo, Wagner, & West, 2004). Patients obtain appointments through an appointment scheduling system which operates based on the scheduling model in the specific clinic setting. Scheduling model and clinic setting context are the two major elements of the patient appointment scheduling system.

There are three major scheduling models: centralized scheduling model (CSM), decentralized scheduling model (DSM) and hybrid scheduling model (HSM). The major difference between the three scheduling models is the number of schedulers that should be involved to schedule all the requested appointments for a patient.

In CSM, patients only contact one scheduler for all the requested appointments. There are two configurations for CSM. In the first configuration, the schedulers are located in a centralized department and the incoming requests are directed to this department. In the second configuration, the schedulers are distributed between individual clinics and the clinic setting. Both of the configurations have centralized scheduling software but the complexity is different.

In DSM, the patients should call schedulers in individual clinics for the requested multiple appointments. Finally, HSM is the combination of CSM and DSM. HSM forms different clusters and assigns single or multiple clinics to each cluster. All the clinics in a same cluster schedule appointments based on CSM, but different clusters are like individual clinics with DSM decision structure.

Clinic settings' context includes but is not limited to the randomness in arrival of requested appointments, proportion of multiple requested appointments, randomness of pattern for the requested multiple appointments, requested appointments type and randomness in the proportion of appointment types.

Indicators of efficiency include but are not limited to patient satisfaction, resource utilization and implementation cost. Patient satisfaction is a qualitative indicator and can be measured in different aspect. Accessibility is an important determinant of patient satisfaction (Gupta & Denton, 2008). Accessibility can be measured through average waiting time before connecting to a scheduler and call duration for getting the requested appointments. Resource utilization can be measured through schedulers' utilization and implementation cost can be measured through scheduling software cost, schedulers' training cost and number of schedulers.

When there is not an appropriate match between the context of the clinic setting and the scheduling model in the setting, the appointment scheduling system would operate inefficiently. In other word, inefficiency in appointment scheduling system reflects a lack of balance between indicators of efficiency.

For example, in the presence of considerable proportion of multiple appointments, DSM scheduling system would be inefficient, because accessibility and resource utilization would be quite low, although implementation cost is economical yet no balance would be found between indicators of efficiency.

In the presence of low proportion of multiple appointments (below 10 percent), CSM scheduling system would be inefficient, because accessibility and resource utilization would be highly provided in the presence of huge implementation cost and no balance would be found between indicators of efficiency.

In the presence of CSM for the clinic setting with high proportion of single appointments, the high cost of advanced scheduling software and schedulers' training is imposed to the hospital, because there is no need for sharing information and coordinating between clinics. For the clinic settings with high proportion of multiple appointments, DSM cause high patient dissatisfaction. As a result, the significance of selecting patient appointment scheduling model in terms of clinic setting context is to provide a balance between patient satisfaction and clinic setting's cost.

The criticality of providing efficient patient appointment scheduling system makes selecting the most efficient scheduling model necessary. The developed procedure in this study determines the most efficient scheduling model for a specific clinic setting and analyzing the inefficient configurations based on the integration of simulation and data envelopment analysis (DEA).

1.2. Research objective and methodology

The objective of this study is to develop a procedure to provide an efficient operation of appointment scheduling system. The developed procedure provides answers for the following three questions:

- Which configurations of the scheduling models are efficient for a specific clinic setting?
- How can the decision makers analyze the inefficient configurations for a specific clinic setting and improve them?

- How can the decision makers select the most efficient configuration of the scheduling model for a specific clinic setting?

The hierarchy for the developed procedure is data collection to estimate distribution for demand arrival and service time, determine the important factors' values for the clinic setting, generate scheduling models' configurations, design simulation models for the generated configurations, assign the simulation output to DEA output, collect DEA inputs, select the question and run determined DEA model to answer the selected question.

DEA has three main elements which are decision making units (DMU), inputs and outputs. DMU is the set of units that DEA is applied for comparing them. The DMUs in this study are different configurations of scheduling models.

DEA approach defines some indicators to evaluate the relative efficiency between different configurations. The indicators that should be maximized are DEA outputs and the indicators that should be minimized are DEA inputs. There are two basic DEA models which are Charnes-Cooper-Rhodes (CCR) and Banker-Charnes-Cooper (BCC) models. The DEA model would be selected based on the question that should be examined.

To examine the first question (Which configurations of the scheduling models are efficient) CCR model is run. To examine the second question (How can the decision makers select the most efficient configuration of the scheduling model), the three terms of inefficiency evaluated are technical inefficiency, scale inefficiency and mix inefficiency. Technical inefficiency presents inefficiency in resource allocation and is obtained by running BCC model. Scale inefficiency presents a need for higher technology and is evaluated by both BCC and CCR models. Finally, mix inefficiency identifies extras in inputs and shortfalls in outputs and is obtained by CCR model. If the configuration is technically efficient, by removing the mix inefficiency, the inefficient configuration can be improved thereby becoming efficient.

To answer the last question (How can the decision makers select the most efficient configuration of the scheduling model), DEA minimax approach that is based on (CCR model) is run.

1.3. Research contribution

This study is conducted to address the significance of the match between the scheduling model and the clinic setting context to provide an efficient patient appointment scheduling system. The clinic setting with high proportion of request for single appointment and CSM scheduling model impose a high cost to the setting and the clinic setting with high proportion of request for multiple appointments and DSM scheduling model makes patient dissatisfaction.

This research contributed to the methodology for selecting the most efficient patient scheduling model for a specific clinic setting, as well as analyzing the inefficient scheduling models and identifying parameters that may be altered to improve the overall system efficiency. Integration of simulation and DEA is the developed methodology to obtain the determined objectives. DEA is commonly used to compare the efficiency (or productivity) of various DMU's across a cross-section of organizations, or across a finite time-horizon. By integrating simulation scenarios and DEA, this research has improved the state-of-the-art by allowing decision makers to evaluate the efficiency of proposed new scenarios against existing settings (both before and after implementation).

The developed methodology can also be extended to other areas of the service and manufacturing industry. The examples for service industry are banks and hotels. In banking industry, the important factor for the bank setting should be identified, these include but are not limited to classifying the type of customers as well as their transaction request types (e.g. Deposit, Withdrawal, etc) and evaluating what should be the skill level of the bank tellers and operators or how many tasks should be assigned to different staff to provide operational efficiency and maximize customer satisfaction. Different scenarios for staff allocation would be modeled by simulation and the efficiency of different scenarios would be evaluated through DEA.

In hotel industry, the examples for the important factors in the hotel setting are different types of guests and their room request types (e.g. Single, Double, Suite, etc). The objective would be determining the number of hotel rooms with specific capacity in the design stage. Different scenarios for customer demand and room allocation are generated by simulation and the most efficient scenario would be determined by DEA.

The examples of manufacturing industry are facility layout design and cellular manufacturing system (CMS). The most efficient facility layout and the most efficient cell formation and operator allocation in CSM would be determined by the integration of simulation and DEA.

The structure of this study is as follow: Section 2 presents a literature review for the different scheduling models, data envelopment analysis and simulation analysis in hospitals/health clinics. Section 3, describes the problem statement as well as the framework structure required to come to identify the most efficient configuration. Section

4, presents the proposed methodology that is used for determining the most efficient scheduling model. Section 5 presents a case study in a local hospital for finding the reliability of the developed framework. Finally the conclusions for the developed study and other useful methodology that can be integrated with DEA in future research (for evaluating the efficiency of patient appointment scheduling system) are proposed.

CHAPTER 2. LITERATURE REVIEW

2.1. Scheduling models

Patient satisfaction has the highest priority for the efficiency of health care system (Vermeulen, Bohte, Elkhuizen, Bakker, & Poutre, 2008). Long waiting time is one of the major reasons for patient's dissatisfaction (Vissers, 1998). To address the patients' long waiting time for getting appointment as well as long cycle time for being scheduled, a consistent patient appointment scheduling model with the setting properties should be developed (Nealon & Moreno, 2003).

An efficient appointment scheduling model should be flexible to the demand variation in the system. Demand variation can lead to bottleneck that is one of the main reason of patient's long waiting time (Vermeulen, Bohte, Elkhuizen, Bakker, & Poutre, 2008). Demand variation leads to inefficient utilization of the mostly shared resources. Diagnostic equipments has degree of sharing because diagnostic test is one of the crucial steps in treatment of many patients group and allocating a fixed capacity to each patients group leads to underutilization of the resource because of the inherent variation in the system. If the resource allocation would be flexible with demand variation the resource would be utilized efficiently (Vermeulen, Bohte, Elkhuizen, Bakker, & Poutre, 2008).

There are three major scheduling models, centralized scheduling model (CSM), decentralized Scheduling model (DSM) and hybrid scheduling model (HSM). The followings sub-section explains the three major scheduling models along with their advantages and disadvantages.

2.1.1. Centralized scheduling model

In centralized scheduling model (CSM), the schedulers have the ability to schedule any clinic within the clinic setting. CSM is important, when the interaction levels between the clinics are high. Centralized scheduling can be done by either having a designated space for centralized scheduling department or through the advanced scheduling software that schedulers in individual clinics are able to schedule appointments for the rest of clinics in the setting. The decision structure for CSM is shown in Figure 2.1 (Zhang, Gonela, & Aslani, 2011).

Centralized scheduling model provides greater uniformity in how appointments are handled and better ability to monitor the entire process. From the patient's perspective, a centralized appointment model allows the patient to use one contact to achieve multiple appointments with different providers and services (Hooten, 1990).

On the other hand, there may be redundancy in a centralized model and waste of resources that was never intended (Hooten, 1990). In centralized scheduling model, the clerks must be trained to do multi-tasking which will enable schedulers to easily switch between tasks during peak hours. This is the most desirable model for scheduling appointments. However in the case of unavailability of the required support for implementing the CSM or if the interaction level between clinics is not high, other scheduling models would be able to provide the acceptable service level.

The appointment scheduling system for the ancillary clinics with diagnostic services that is necessary in the treatment process of other clinics is not efficient in the decentralized scheduling system. The centralized scheduling system is efficient for these ancillary clinics because the patients are scheduled for lab tests by having information about the resource availability in real time basis.

For having an efficient centralized scheduling system, there should be an efficient staffing level in the department to handle the problem of patients' long waiting time to be connected to schedulers as well as long cycle time to be scheduled for requested appointments. Queuing theory is an appropriate approach to decide the required number of staffs for the subjective service level (Agnihotri & Taylor, 1991).

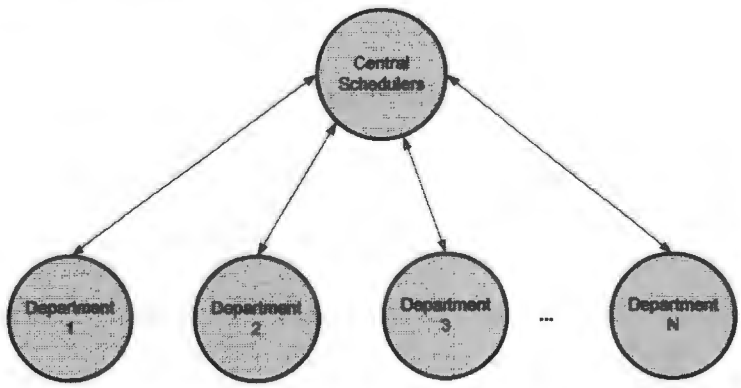

Figure 2.1. Decision process structure in CSM. Adapted from "Development of Centralized, Decentralized and Hybrid Scheduling Model" by Zhang, Gonela, & Aslani, 2011.

There are similarities between centralized scheduling and flow shop manufacturing. Selecting the flow shop manufacturing is useful when there is a demand for high volume products with low variability. Centralized scheduling is useful when the number of clinics is high and there is high interaction level between majorities of the clinics in the hospital (high level). Presence of high interaction level can be interpreted as high rate of multiple appointment demand.

Advantages and disadvantages of CSM for appointment scheduling phase are as follow:

Advantages:

1. One point of contact for patients with multiple appointments that leads to reduced patient's waiting time for getting an appointment and less total registration time.
2. Considerable reduction in rescheduling rate because of appropriate information sharing and coordination between ancillary departments and the referring department.
3. Reduction in total cycle time to be scheduled and patient's waiting time before connecting to a scheduler because of the accessibility of the required information through the shared information system.
4. Enhancing the utilization of resources (nurse, equipment and space) through sharing.
5. Considerable reduction in the conflict between treatment because of the availability of temporal constraints through the sharing information system (Marinagi, Spyropoulos, Papatheodorou, & Kokkotos, 2000).

Disadvantages:

1. The control power of scheduling is mostly in the hand of schedulers and not providers that make them dissatisfied.
2. In the CSM configuration, schedulers are only centralized and there are not individual schedulers for individual departments. The scheduling errors would be increased because the schedulers are unaware of the process happening in each clinic.

2.1.2. Decentralized scheduling model

This model uses individual schedulers for scheduling different clinics in the setting. The advantages of this model are: sufficient experience of each scheduler that leads to fewer numbers of errors in scheduling, and higher acceptability of walk-in patients. However implementation of this model is useful in setting with small number of clinics (like standalone setting or settings with negligible interaction level).

There are similarities between this model and job shop manufacturing. In job shop manufacturing the same type machines are grouped together which is similar to each scheduler just serving the patients for specific clinics. The decision structure for DSM is shown in Figure 2.2 (Zhang, Gonela, & Aslani, 2011).

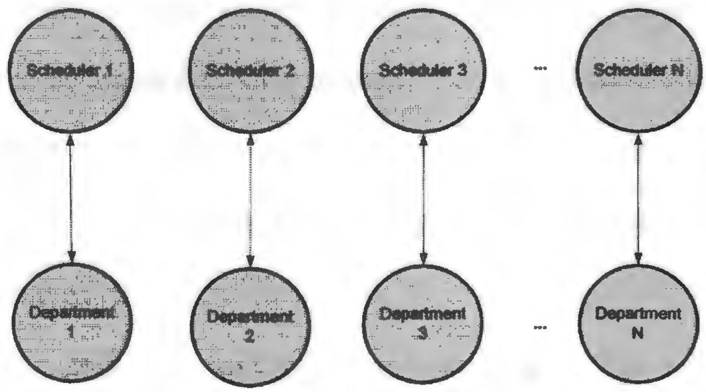

Figure 2.2. Decision process structure in DSM. Adapted from "Development of Centralized, Decentralized and Hybrid Scheduling Model" by Zhang, Gonela, & Aslani, 2011.

Berry & Phanthasomchit (2000) presents some advantages and disadvantages of DSM which are as follow:

Advantages:

1. No change in the current number of staff.

2. Capability for ad hoc scheduling because of the presence of flexibility in the system.
3. Availability of sufficient information.

Disadvantages:

1. Multiple contacts with different schedulers for patients with multiple appointment.
2. Lack of an information sharing system.
3. Underutilization of resources.
4. High number of patients' appointment rescheduling and cancelation.
5. Possibility of conflicts in patient treatment.

2.1.3. Hybrid scheduling model

This model is the combination of centralized and decentralized scheduling models. In this case, some clinics use CSM and the rest use DSM. Clustering the clinics with CSM or DSM is based on the interaction level between clinics. In the presence of setting with high interaction level between clinics and unavailability of advanced scheduling software, HSM is selected instead of CSM. The decision structure for HSM is shown in Figure 2.3 (Zhang, Gonela, & Aslani, 2011).

This model is comparable with cellular manufacturing (CM). In CM, clustering is based on the family parts. Family parts are the products with the same processing requirements. In hybrid scheduling model the clinics with high interaction level are placed in the same cluster.

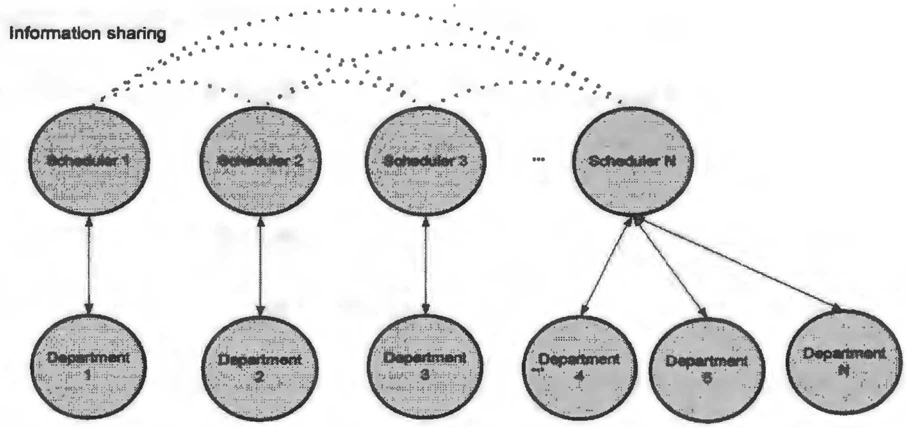

Figure 2.3. Decision process structure in HSM. Adapted from "Development of Centralized, Decentralized and Hybrid Scheduling Model" by Zhang, Gonela, & Aslani, 2011.

2.2. Data envelopment analysis

Data Envelopment Analysis (DEA) is a proper tool to identify relative efficiency between different organizations or different units in one organization in the presence of multiple inputs and outputs with complex relationships (Cooper, Lawrence, & Tone, 2006).

2.2.1. What is DEA?

DEA is an efficiency benchmarking tool from the efficiency frontier family and can determine the relative efficiency of different profitable and non-profitable organizations or different configurations in an organization from the perspectives of layout design, operation, scheduling or other properties within an organization. Different organizations or different configurations of an organization that are evaluated by DEA are considered as Decision Unit (DU). DEA is a member of non-parametric techniques and it can find the relative efficiency of a DU with multiple inputs and outputs that can be incommensurate. DEA discriminates between different inputs and outputs by assigning different weight to each of them. The efficiency of a DU with multiple inputs and outputs in DEA is calculated from the ratio of sum of weighted outputs to the sum of weighted inputs of a DU (Avkiran, 2001).

DEA models can be categorized into input-oriented and output-oriented models. In the input oriented model, input is minimized to get the same output, however in the output oriented models the output is maximized for the same input (Ramanathan, 2003). Generally DEA methodology has two main properties. The first property is to provide a reference for each inefficient DU as a target to find out the changes that can shift an inefficient DU into an efficient one (Cooper, Lawrence, & Tone, 2006).

Two different kinds of efficiency can be determined through DEA, weak efficiency (Farrell efficiency) and strong efficiency (Pareto). The constraint of weak efficiency is that an increase in output should not lead to an increase in input. Strong efficiency has one more constraint to weak efficiency that makes the strong efficiency constraint stricter. This additional property of strong efficiency is that an increase in output should not lead to a decrease in other output (Avkiran, 2001).

There are some advantages and disadvantages for DEA method. The advantage of DEA is, the performance of each unit is based on unit's performance that would be presented through piecewise linear programming not base on some presumptions (Avkiran, 2001).

The following are the disadvantages of DEA method: firstly, the efficiency of each unit is valid through that sample test and not between the data outside of sample test (Avkiran, 2001). Secondly, if the numbers of compared DUs are not sufficiently larger than the total number of inputs and outputs, then too many DUs would be identified as efficient units. Identifying too many DUs as efficient is the evidence of local optimum trap (Li & Reeves, 1999). The third drawback of DEA method is the situation in which a determined efficient DU assignes too small weight for an input or/and too large weight for an output which cannot be applied in reality. (Li & Reeves, 1999). Some multiple criteria approaches have been developed through some studies that can eliminate the effects of the thrid drawback such as minimizing the inefficiency measure of a evaluated decision unit, minimizing the sum of the inefficiency measures of all the decision units or minimizing the maximum inefficienct measure between all the DUs (Li & Reeves, 1999).

The performance model in DEA is defines based on the organization's objectives, so DEA takes objectives of the organization as outputs and organization resources as inputs. Basically inputs in DEA are the factors that should be minimized and outputs are the factors that should be maximized (Ramanathan, 2003).

The basic DEA model to discriminate between efficient and inefficient configurations is obtained from the study of Oral & Yolalan (1990) and is presented in the following:

$$\text{Maximize } \sum_{r=1}^{S} u_r \, y_{rB} / \sum_{i=1}^{m} v_i \, x_{iB} \tag{2.1}$$

$$\text{Subject to } \sum_{r=1}^{S} u_r \, y_{rj} / \sum_{i=1}^{m} v_i \, x_{ij} \leq 1 \qquad \forall j$$

(2.2)

$$j = 1,2,\ldots,n$$

$$u_r \, v_i \geq \epsilon > 0$$

$$r = 1,\ldots,S, \; i=1,\ldots,m$$

where

i	Index number of inputs	i=1, 2,..,m
r	Index number of outputs	r=1,2,…,s
j	Index number of decision units	j=1,2,…,N

y_{rj} = quantity of output r for the decision unit j

x_{ij} = quantity of input i for the decision unit j

u_r = the weight given to output r

v_i = the weight given to input i

ϵ = small positive number

DEA is determining efficiency by defining the fraction that is sum of weighted outputs over the sume of weighted inputs for a particular DU (Equation 2.1). Equation 2.2 demonstarte the constarint to make DEA efficieny score less than or equal to one (Oral & Yolalan, 1990).

Charnes, Cooper and Rhods develop a model to transfer the non-linear basic DEA to the linear model that is known as CCR. One of the properties of CCR model is its capability for identifying extra in inputs and outputs' shortfall for providing improvement. This property is called global efficiency (Cooper, Lawrence, & Tone , 2006).

2.2.2. Application of DEA

DEA is a non-parametric approach for finding the relative efficiency of either different organizations or different units in one organization.DEA is applied in the literatures as a benchmarking tool with the property of integrating performance evaluation with decision making through finding the relative efficiency of the selected units in an organization and make decision for the required modifications (Oral & Yolalan, 1990).

The applications of DEA in manufacturing and education area are observed through the literature review. DEA in manufacturing system area is used for evaluating the relative efficiency of facility layout and technology for manufacturing. Ertay, Ruan, & Tuzkaya (2006) and Yang & Kuo (2003) solve the multi objective layout design problem by applying DEA methodology. Layout design problem is considered as a multi objective problem for considering an efficient layout for which both quantitative and qualitive criteria should be considered. DEA methodology can consider both quantitative and qualitative data simultanously so it is recognized as an appropriate tool for layout design problem.

Ertay, Ruan, & Tuzkaya (2006) developed a framework with the contribution of both analytical hierarchy process (AHP) and data envelopment analysis (DEA) for detemining an efficient layout in manufacturing system. Ertay, Ruan, & Tuzkaya (2006) categorizes the quantative data into exact and vague. Material handling cost is a exact data which is collected directly, while adjacency score is a vague data that is obtained through fuzzy set theroy and AHP is used for finding qualitative data like Flexibility and Quality. The Material handling cost and adjacency score that should be minimized are considered as inputs for DEA model and the data that should be maximized are considered as output which are shape ratio, flexibility, quality and hand-carry utility.

Ertay, Ruan, & Tuzkaya (2006) use classical output-oriented CCR to minimize the inefficienct measure and multiple layout alternatives are found to be efficient. In order to narrow down the solution space, the authors modify the objective of classical output-oriented CCR to minimizing the maximum inefficiency measure (minimax efficiency).

Yang & Kuo (2003) solve the layout design problem with the contibution of AHP and DEA methodologies in the presence of both quantitative and qualititative data. The quantitive data are distance, adjacency and shape ratio. The qualitative data are flexibility, accessibility and maintenance which are obtained by AHP method. All of the mentioned quantitative and qualitative data are DEA model outputs and layout related cost which is constant in the DEA model inputs.

The DEA model that is used for selecting the efficient layout is an output-oriented BCC model with a single constant input which is the same as output-oriented BCC model without inputs (Yang & Kuo, 2003).

For selecting the technologies with high complexity and variation like robot selection in manufacturing system, DEA method is an appropriate tool because the defined performance measures for the robot are dependent and highly complex. DEA methodology considers the interrelationships between inputs and outputs so it would be a proper tool for technology selection (Khouja, 1995).

Khouja (1995) develops a research methodology for selecting the efficient robot for a manufacturing system by the integration of DEA and Multi Attribute Decision Making (MADM) methods. This study includes two phases. In phase one, the technologies that hasve good fit with the manufacturing system would be selected by DEA and in the second phase, the best technology from the selected technologies in phase one would be determined through MADM. The inputs for DEA are cost and repeatability and the outputs are speed and load capacity (Khouja, 1995).

Cellular manufacturing (CM) is from the group technology family and cell formation is the major procedure to come with the efficient layout in CM. (Shafer & Bradford, 1995). DEA is an appropriate tool for dealing with multiple layouts that are produced with the cell formation procedure and selecting the efficient layouts. The dual of the classical CCR is used as the DEA methodology that considers number of clusters and number of machines as inputs and average work-in process level, average flow time and average worker utilization as outputs. The considered outputs are generated by simulation (Shafer & Bradford, 1995).

There are some articles that have used DEA for determining the efficiency of departments in a university. The inputs for these researches are: number of academic and non-academic staff, research income, direct expenditures, operating costs and salaries. The outputs are: number of students, research rating grants, publications and contact hours (Avkiran, 2001).There are some articles which have been used DEA for assessing efficiency of different universities. For these articles, the selected inputs are faculty salaries, administrative overhead and total investment in physical plant. The outputs are number of undergraduate and graduate enrollments, total semester credit hours, and federal private research fund (Avkiran, 2001).

Performance models are defined according to the objective that administrators have defined. The performance models that Avkiran (2001) has defined are overall performance model, education delivery of universities model and fee-payment enrollment model.

Avkiran (2001) uses production theory to determine the inputs and outputs for the performance models in his study. For overall performance model the inputs are number of academic and non-academic staffs and outputs are undergraduate, graduate enrollment and research quantum which is research component of federal funds that includes research grants, number of research completion and number of publications. For the delivery of educational services model the inputs are the same as previous model and the outputs are student retention rate, student progress rate and graduate full-time employment. Finally for the performance on fee-paying enrollment the inputs are the same as the other two models and the outputs are overseas fee-paying enrollments and non-overseas fee paying postgraduate enrollment (Avkiran, 2001).

Four approaches can be used for analyzing options in DEA that are: input-oriented, output-oriented, constant return to scale (CRS) and variable return to scale (VRS). For input-oriented method, the reduction in inputs is possible as long as outputs do not drop, this approach is useful when cost saving or downsizing is the goal. For output oriented method the productivity is raised without increasing the resource base. For CRS method, there is not significant relationship between DU size and efficiency, for example small and large universities do the same in converting input to output. For VRS method the rise in input is not consistent to the rise in output. The number of academic staff can be interpreted as the DU size, when the relation between efficiency and DU size would be cleared in large sample through VRS (Avkiran, 2001).

For choosing between CRS and VRS, the performance models should be run under both methods and if there exist a big difference between efficiency scores, VRS would be selected. CRS shows technical efficiency while VRS shows pure technical efficiency. Technical efficiency is decomposed into pure technical efficiency and scale efficiency. VRS methods also categorized into increasing return to scale (IRS) and decreasing return to scale (DRS) (Avkiran, 2001).

Kirigia, Emrouznejad, & Sambol (2002) study the efficiency of public hospitals in Kenya has been by DEA methodology. Technical efficiency and scale efficiency scores have been computed. The technical efficiency score of inefficient hospitals shows the reducing percentage in input utilization and the scale inefficiency score of inefficient hospitals show the possible percentage of increase in outputs. The new policies have been made regarding to excess in inputs (Kirigia, Emrouznejad, & Sambol, 2002).

DEA classifies decision units into efficient and non-efficient units and the achieved efficiency scores in DEA are weights, but because they have been obtained through different comparisons, so in order to compare them together the value of weights should be converted into the same scale. Sinuay & Friedman (1998) have developed a non-linear programming approach to unitize the weights and rank them.

The advantages of Discriminate Data Envelopment Analysis of Ratio (DR/DEA) that is proposed by Sinuay & Friedman (1998) are decision units that are previously ranked into efficient and non-efficient units by DEA and their common weights are obtained through non-linear optimization of goodness of separation between efficient and non-efficient units are fully ranked. Secondly, make separation between inputs and outputs and finally, the fit between DEA and DR/DEA can be validated through non-parametric statistical test.

DEA is a good tool to be used in the evaluation stage of Facility Layout Design (FLD). Ertay, Ruan, & Tuzkaya (2006) apply DEA for determining the efficiency of 19 facility layout alternatives and considering both qualitative and quantitative data at the same time. Quantitative data are considered as inputs and adjacency score present activity relationship is computed through fuzzy method. Qualitative data are collected through AHP method and considered as outputs. The inputs are adjacency scores, shape ratios, material handling cost and material handling vehicle utilization. The outputs are volume flexibility, variety flexibility, production quality and product quality.

The CCR model is used to change the fractional program to a linear program and the BCC model without input because the cost of layout design stage is not considerable is a good approach for layout performance frontier problem. For considering different demand scenarios and minimizing the material handling cost the robust layout should be designed. Classical DEA does not choose the robust alternative, as DEA calculates the efficiency score of each alternative separately, so for getting the robust alternative, minimax efficiency is used.

Application of DEA in operation efficiency of bank branches is studied by Oral & Yolalan (1990). Operation efficiency of banks used to be measured by some classical approaches like financial ratio which is mostly effective in a short run and evaluate all operation, marketing and financing in one set. DEA covers the deficiency of the traditional approaches. Two DEA models are used in the study of Oral & Yolalan (1990). In the first DEA model, the efficiency of bank branches are investigated individually and then by continuing to the second model and defining the most frequent member in efficient reference set as the global leader, the branches can be compared together to reallocate the resources. Two performance models have been defined, serviceability model and profitability model because the objective is attracting more clients to the bank by having higher serviceability and it should provide profitability as well. The inputs for serviceability model are number of personnel, number of terminals, number of commercial accounts, number of saving accounts and number of credit applications , and the outputs are time spent on general services , time on credits, time on deposits and time on foreign exchange. Five combinations of these inputs and outputs are used to get more valid results from the perspective of the branch manager. Profitability model inputs are personnel expenses, administrative expenses, depreciation, interest paid and the outputs are interests earned and non-interest income. Three combinations of these inputs and outputs are generated.

One of the applications of DEA is evaluation the performance of an organization and performance evaluation is a major element for planning and construction policies. Chiang (2006) evaluates the performance of 25 hotels in Thailand. Thailand is a country with lots of tourism attraction so hotel industry would act as a complementary element that encourages more tourists to travel. There is an intense competition between hotels in Thailand. In order to win this competition the hotels should perform well compare to others and DEA is used to identify resource overutilization and make the hotel performance efficient by reallocation of resources. Evaluation should be done over the homogenous data that have the same geographic characteristics. Chiang (2006) get the data for his study from Annual Operation Report of the international Tourist Hotel. The inputs for the study is number of hotel rooms, capacity that is dedicated to food department, total number of employee and total operating cost of the hotel. The outputs are Yielding index that represent daily occupancy and availability of a room, revenue from food, revenue from sources other than rooms and food. Chiang (2006) use CCR for getting overall performance efficiency and BCC for getting pure technical efficiency so it would be clear the efficiency comes from the shortage in technical efficiency or scale efficiency.

DEA is applied for identifying the existence of congestion and separate it from technical inefficiency. Technical inefficiency happens when improving in outputs and inputs happen without changing other inputs and outputs, but congestion happen when decreasing one or more input make one or more output increase simultaneously (Cooper, Deng, Gu, Li, & Thrall, 2001). After implementing the "iron rice bowl" policy in Chinese industry, many industries went bankrupt because of appearance of congestion, so Chinese government decided on a massive layoff for preventing these problems, but massive layoff generates social tension and Cooper, Deng, Gu, Li, & Thrall (2001) study how the inefficiency can be managed without removing the congestion. Two stage BCC model from DEA methodology is selected. The first stage is a radial measure model for computing technical inefficiency that decomposed to pure technical inefficiency and mix inefficiency. In the second stage an additive model is developed for measuring the congestion. Finally a model is designed for determining how the output can be improved without removing the congestion because removing of that congesting element would lead to some tension in that organization or society. Cooper, Deng, Gu, Li, & Thrall (2001) have presented his study in automobile and textile industry and the congestion in labor is more than the congestion in capital, but it is observed that the efficient management over capital can have the same result as labor reduction so improvement can be made beside the existence of congestion in the system.

Previously, in the healthcare area different studies are done that use DEA for the evaluation stage. However some of them have compared some criteria for different hospitals not in one hospital. In the other areas of study like facility layout design a study has been done for evaluating different layouts for a facility by using DEA (Ertay, Ruan, & Tuzkaya, 2006). A study has been done for evaluating different technologies by DEA (Khouja, 1995).

CHAPTER 3. PROBLEM STATEMENT AND SYSTEM CONFIGURATION

This research is conducted to address the problem of assigning most efficient scheduling model to the specific clinic setting. Appointment scheduling system assign clinics' time slots to the requested appointments (Guo, Wagner, & West, 2004). There are three different appointment scheduling models: centralized scheduling model (CSM), decentralized scheduling model (DSM) and hybrid scheduling model (HSM).

In a clinic setting with CSM, all the requested appointments for a patient would be scheduled by one scheduler, in other words the requested appointments would be scheduled through one call to one scheduler. CSM can either have the schedulers who are centralized in scheduling department or the schedulers who are distributed between individual clinics and more advanced scheduling software to provide coordination between schedulers in different clinics.

In a clinic setting with DSM, the requested appointments for a patient should be broken down into single requests. The patient should contacts separately to individual schedulers in charge of different clinics to be scheduled for all the single requested appointments.

In a clinic setting with HSM, some clinics have CSM scheduling model and some clinics have DSM scheduling model. The requested appointments should be decomposed into the clinics with CSM and clinics with DSM.

Clinic settings are different in terms of several important factors such as: arrival of requests for appointments, proportion of multiple appointments versus single appointments, randomness in the sequence of requested multiple appointments in the stage of appointment scheduling and no-shows, scheduling rules, patients' and providers' punctuality in the stage of visiting the clinic settings.

Efficiency is evaluated through three categories of indicators relevant to appointment scheduling systems: satisfaction indicator for patients, staffs and managers, resource utilization indictor for schedulers, space and equipments and cost indicator for scheduling software and schedulers' training.

Inefficiency in the operation of scheduling system presents the lack of balance between indicators of efficiency. The balance is provided if the scheduling model is selected based on the important factors of clinics setting.

Multiple appointments are considered as one of the major problems in appointment scheduling models (Nealon & Moreno, 2003). For example, if the proportion of multiple appointments in a setting with CSM is inconsiderable, the scheduling system operates inefficiently because the balance between the three indicators of efficiency is not provided (too large value for cost indicators besides high value for indicators of satisfaction and resource utilization do not provide balance because other scheduling models with lower cost indicators provides acceptable value for indicators of satisfaction and resource utilization as well).

This study considers arrival of the requested appointments and proportion of multiple appointments as the important factors in clinic settings. Interaction index which is defined as the proportion of multiple requested appointments over total requested appointments determines the difference between clinic settings. If the interaction ratio is less than 15 percent, the interaction level in the setting is low. If the interaction ratio is from 25 percent to 50 percent, the interaction level in the setting is medium and finally if the interaction ratio is more than 70 percent the interaction level in the setting is high.

A procedure is developed in terms of the integration of simulation and Data Envelopment Analysis to address the problem of assigning most efficient scheduling model to the specific clinic setting. The proposed methodology compares different configuration of scheduling models based on the three categories of efficiency indicators.

In this study, satisfaction indicator is measured with patient satisfaction. Resource utilization indicator is measured with schedulers' utilization and number of schedulers. Cost indicator is measured with scheduling software cost and schedulers' training cost. Patient satisfaction is a qualitative indicator and it can be estimated with quantitative factors. Accessibility is a valid estimate for patient satisfaction (Gupta & Denton, 2008). In this study, accessibility is measured by average waiting time before connecting to a scheduler and average call duration.

In order to select the most efficient scheduling model for a clinic setting, DEA approach compares different decision making units which are generated from configurations of scheduling models by using the efficiency indicators that should be minimized as DEA inputs and efficiency indicators that should be maximized as DEA outputs. DEA is an appropriate decision tool for the multi-objective settings with many inputs and outputs that have complex relationship. The perspective of this research is to determine the following things for a specific clinic setting:

- Which configurations of the scheduling models are efficient?

- How can the decision makers analyze the inefficient configurations and improve them?
- How can the decision makers select the most efficient configuration of the scheduling model?

CHAPTER 4. PROPOSED METHODOLOGY

Three major patient appointment scheduling models are developed based on the mutual impacts of clinic flows and patient appointment scheduling. Efficiency of different scheduling models varies for different types of settings. Three different types of settings are defined in terms of percentage of multiple requested appointments versus single requested appointments.

Figure 4.1 illustrates the procedure for determining the most efficient scheduling model. The developed procedure illustrates the following 5 general steps to select the most efficient configuration of the patient appointment scheduling model for a specific type of clinic setting.

1. Data collection
2. Identify clinic setting
3. Configuration generation
4. Simulation
5. DEA approach

4.1. Clinic settings

The focus of data collection in this study is to provide numerical values for the determined inputs and outputs of the DEA model. The factors that should be minimized are the inputs of DEA and the factors that should be maximized are the outputs of DEA.

The DEA inputs are scheduling software cost, training cost and number of schedulers. The two cost-based inputs are obtained from the hospital administrator and vendors and number of schedulers is obtained from the queuing theory formula for the subjective service level.

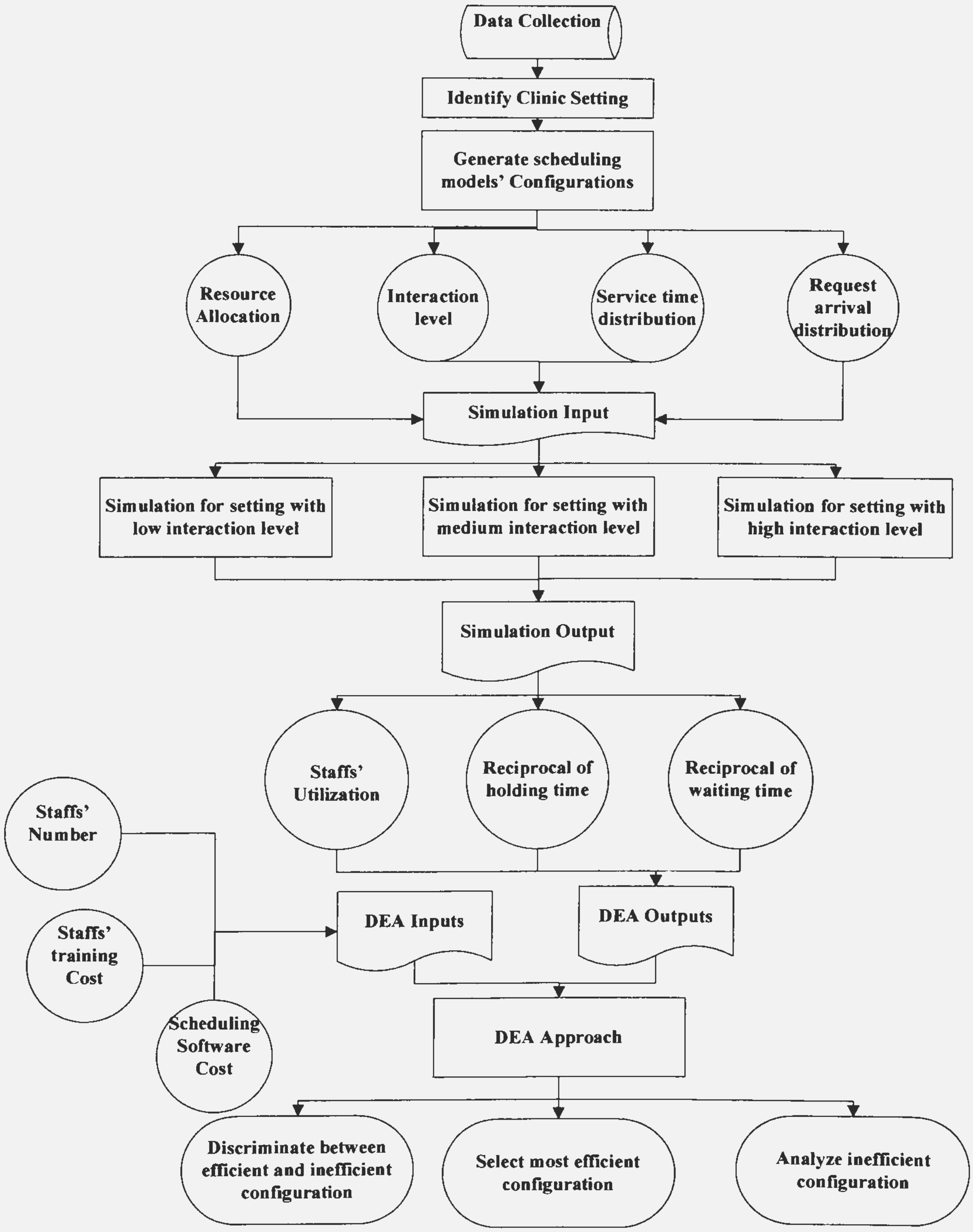

Figure 4.1. Framework for determining the most efficient scheduling model.

The DEA outputs are obtained by conducting a simulation study. The simulation inputs are obtained in the data collection stage. The simulation inputs include stochastic characteristics of the setting which are percentage of multiple appointments versus single appointments, percentage of different types of appointments, the pattern of request arrival for different number of clinics and for all the three appointment type, the scheduling time pattern for the schedulers with different skill level, the service level and the required number of the schedulers for different configuration.

The percentage of multiple appointments is obtained from the available historical data in the database. The clinics' interaction index is the ratio of total number of multiple appointments over total number of appointments.

If the interaction index is less than 15 percent, the interaction level in the setting is low. If the interaction index is from 25 percent to 50 percent, the interaction level in the setting is medium and finally if the interaction index is more than 70 percent the interaction level in the setting is high. The interval from 15 percent to 25 percent is the threshold interval and the setting can be considered as either low or medium interaction level. In addition, the interval from 50 percent to 70 percent is the threshold interval and the setting can be considered as either medium or high interaction level.

4.2. Configuration generation

Generally there are three major scheduling models: centralized scheduling model (CSM), decentralized scheduling model (DSM) and hybrid scheduling model (HSM). Different configurations for the three major scheduling models are generated to provide sufficient number of alternatives. Sufficient number of alternatives should be evaluated to identify the best configuration of scheduling model for a clinic setting.

4.2.1. Centralized scheduling model's configurations

Two configurations for centralized scheduling model (CSM) are generated in this study: CSM (1) and CSM (2). In CSM (1), the schedulers are single task and they are centralized in the scheduling department. In CSM(2), the schedulers are multi-task and they are distributed between individual clinics.

4.2.1.1. Configuration 1 of CSM: with centralized scheduling department

CSM (1) is presented in Figure 4.2 that is obtained from the study of Zhang, Gonela, & Aslani (2011). The conditions for implementing CSM (1) is demenostrated in Table 4.1.

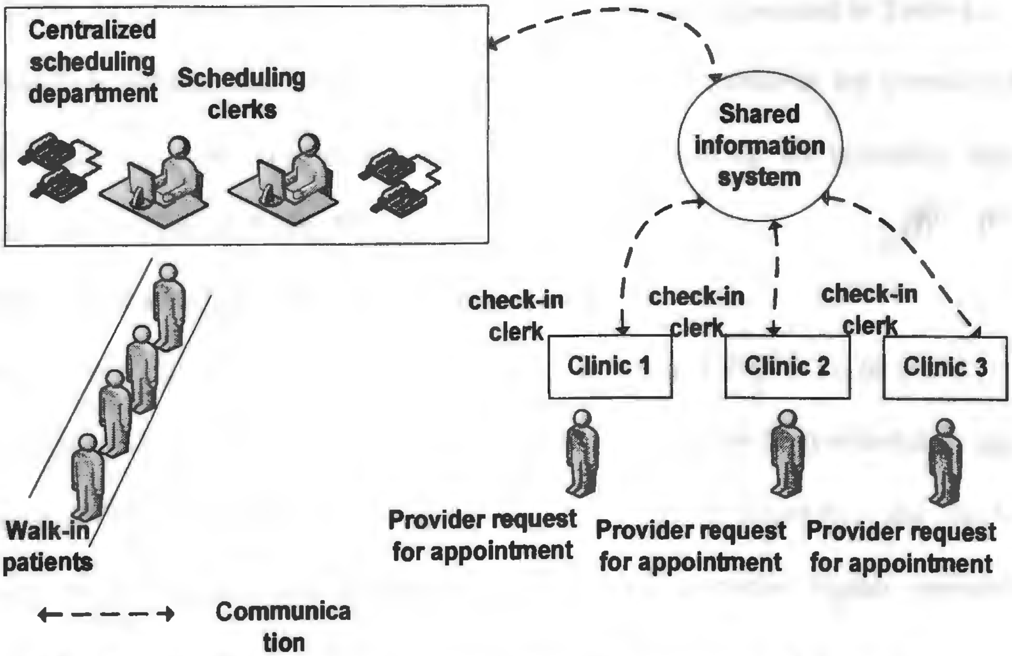

Figure 4.2. Configuration 1 for CSM. Adapted from "Development of Centralized, Decentralized and Hybrid Scheduling Model" by Zhang, Gonela, & Aslani, 2011.

Table 4.1. Conditions for implementation of CSM (1).

Requirements	Condition
Space Requirement	When enough space is available for establishment of scheduling department and check-in space is available at clinics
Information Technology	When information technology is powerful and advanced. Requires centralized information sharing
Personnel requirements	Specialized schedulers capable of handling sophisticated scheduling activities that ranges from single appointment to multiple appointments which requires specialized trainings
Equipment Requirements	Equipment such as printers, computer, furniture's that are needed to set scheduling department are available easily

4.2.1.2. Configuration 2 of CSM: with multitasking and distributed schedulers

CSM (2) is presented in Figure 4.3 that is obtained from the study of Zhang, Gonela, & Aslani (2011). The conditions for implementing CSM (2) is demenostrated in Table 4.2.

CSM (2) is implemented in situations where no space is available for centralized scheduling department. The schedulers in CSM (2) are multitasking for providing high scheduler utilization (Ertay & Ruan, 2005).

4.2.2. Decentralized scheduling model's configurations

Two configurations for DSM are generated: DSM (1) and DSM (2). In DSM (1), schedulers are single task because of the high volume of activities for both scheduling and check-in tasks. In DSM (2), schedulers are multitasking because the scheduling and check-in tasks are not overlapped and schedulers' multitasking provides higher resource utilization (Ertay & Ruan, 2005).

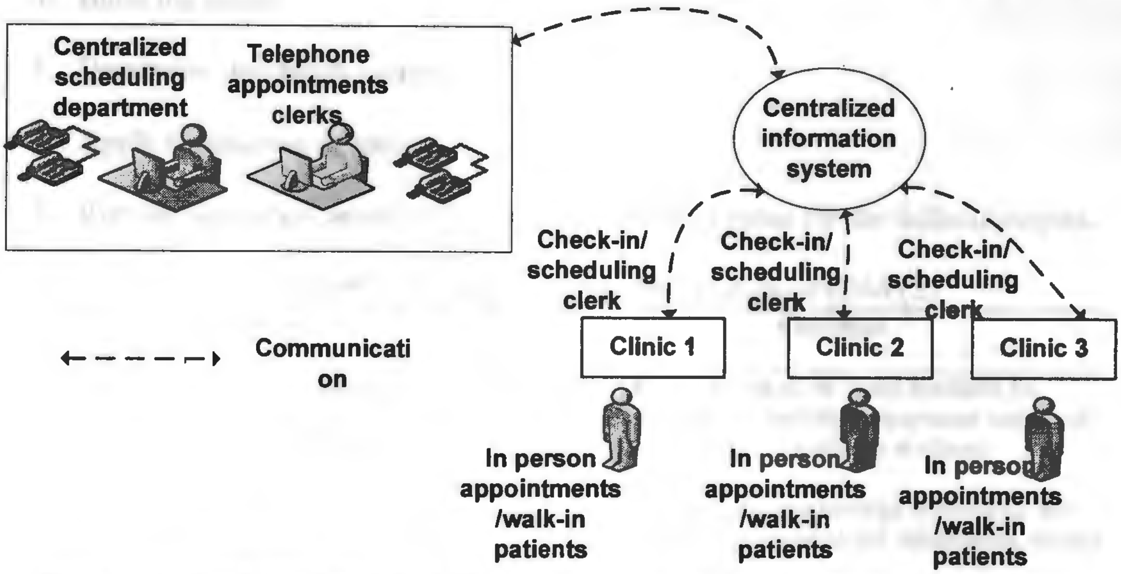

Figure 4.3. Configuration 2 for CSM. Adapted from "Development of Centralized, Decentralized and Hybrid Scheduling Model" by Zhang, Gonela, & Aslani, 2011.

4.2.3. Hybrid scheduling model's configurations

HSM configurations are obtained by defining specific number of clusters for each configuration. Each cluster would include single or multiple clinics.

4.3. Simulation study

The numerical values for the determined DEA outputs in this study are obtained through conducting a simulation study. The simulation study is performed for the three defined clinic settings. The followings are the 5 general steps to conduct the simulation study.

1. Identify feasible number of requested multiple appointments and their probability (Data Collection)
2. Identify the probability of specific requested combination of clinics for all the feasible multiple appointment (Data collection)
3. Collect simulation model inputs

4. Build the model.

5. Determine simulation outputs.

6. Apply termination conditions.

7. Run the simulation model and obtain the numerical value for the defined outputs.

Table 4.2. Conditions for implementation of CSM (2).

Requirements	Condition
Space Requirements	When enough space is not available for establishment of scheduling department and check-in space is available at clinics
Information Technology	When information technology is powerful and advanced. Requires centralized information sharing
Personnel requirements	Clinic schedulers should be capable of handling wide range of scheduling activities that requires specialized training
Equipment Requirements	Designated scheduling area can be obtained at each clinic or check-in clerks can schedule patients resulting in usage of Equipment that are already at these places

4.4. Data envelopment analysis approach

Data Envelopment Analysis (DEA) is applied as an approach in this study to indicate the most efficient configuration as well as the reason for inefficient configurations. The general guideline for the applied DEA approach in this study is illustrated in Figure 4.4.

Section 4.4.1 presents different terms of inefficiency and the interpretation for each term. Section 4.4.2 identifies efficient configurations based on efficiency terms and Section 4.4.3 illustrates how to select the most efficient configuration.

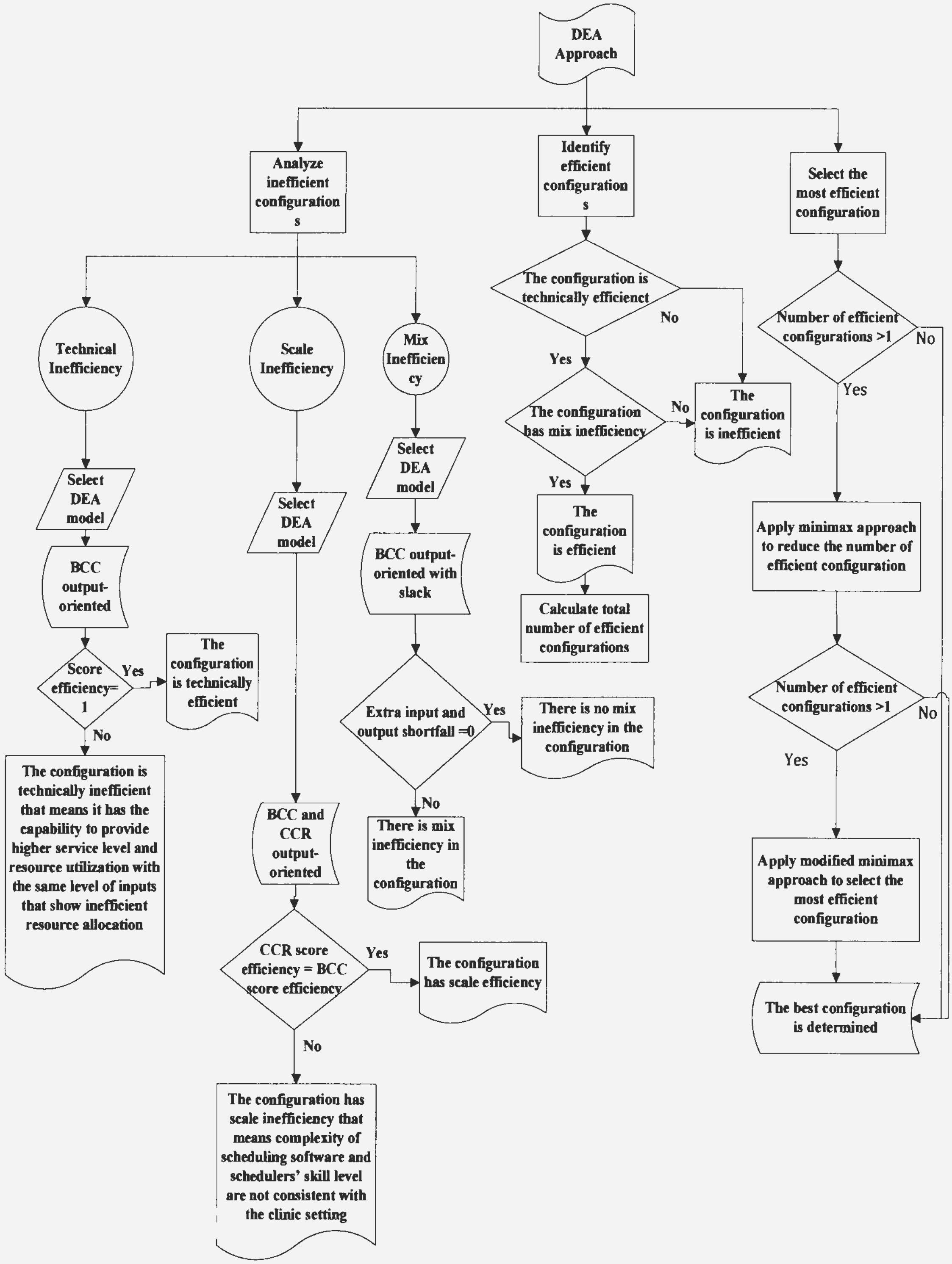

Figure 4.4. DEA approach flowchart.

4.4.1. Analyze inefficient configurations

In this study, three terms of inefficiency are evaluated which are technical inefficiency, scale inefficiency and mix inefficiency. The three following sections illustrate the solutions for the three different inefficiency terms. Table 4.3 illustrates the definition for the three terms of efficiency.

Table 4.3. Different terms of inefficiency.

Inefficiency term	General definition	Interpretation in this study	Applied Model
Technical Inefficiency	The configuration does not use the available capacity efficiently (Venkatesh, 2006 and Ozcan, 2008)	Performance of scheduling software and schedulers do not match their capability.	BCC output-oriented
Scale Inefficiency	the configuration does not use the proper technology with clinic setting's inputs (Venkatesh, 2006 and Ozcan, 2008)	The clinic setting requires more complex scheduling software and higher skill level.	BCC and CCR output-oriented
Mix Inefficiency	There exists extra in inputs and shortfall in outputs	Waiting time and handle time should be decreased and utilization rate should be increased.	CCR output-oriented with slack

4.4.1.1. Technical inefficiency

The models that are introduced by Charnes-Cooper-Rhodes (CCR) and Banker-Charnes-Cooper (BCC) are the major models in DEA. The main difference between CCR and BCC models is that CCR efficiency score does not consider scale inefficiency. As a result BCC efficiency score illustrates pure technical efficiency.

In this study, BCC model is selected to calculate technical efficiency. There are two input-oriented and output-oriented BCC models. BCC output-oriented is selected because the inputs (scheduling software cost, training cost and number of staffs) are constant but the outputs can be changed. Cooper, Lawrence, & Tone (2006) indicates the following formula in their book as BCC output-oriented model.

$$\text{Max } \eta_B \tag{4.1}$$

$$\text{Subject to } \sum_{i=1}^{m} x_{i0} - \sum_{i=1}^{m} \lambda_j\, x_{ij} >= 0 \qquad \forall j \tag{4.2}$$

$$\eta_B \sum_{r=1}^{s} y_{r0} - \sum_{r=1}^{s} \lambda_j\, y_{rj} <= 0 \qquad \forall j \tag{4.3}$$

$$\sum_{j=1}^{N} \lambda_j = 1 \tag{4.4}$$

$$\lambda_j \geq 0$$

where

β :	BCC model	
i	Index number of inputs	i=1, 2,..,m
r	Index number of outputs	r=1,2,…,s
j	Index number of decision units	j=1,2,..,N

η_B = Efficiency Score

y_{rj} = Quantity of output r for the decision unit j

y_{r0} = Quantity of output r for the evaluated decision unit

x_{ij} = Quantity of input i for the decision unit j

x_{i0} = Quantity of input i for the evaluated decision unit

λ_j = The benchmark index for the decision unit j

Equation (4.1) is the objective function to maximize η_B for increasing y_{r0} to $\eta_B y_{r0}$ without increasing x_{i0}. Equation (4.2) and Equation (4.3) confirm that the improved decision unit is located in the feasible region. Equation (4.4) defines the feasible region based on convex hull concept. If η_B is equal to one, the configuration is technically efficient, otherwise the configuration would be inefficient. The benchmark index for the efficient configuration is equal to 1, but for the inefficient configuration is equal to zero. The reference set for inefficient configurations, are the configuration with benchmark index greater than zero.

4.4.1.2. Scale inefficiency

In the first place, the CCR output-oriented model should be run, and its efficiency score would be compared with the obtained BCC efficiency scores for technical inefficiency. If CCR model efficiency score is not equal to the BCC efficiency scores, the configuration has scale inefficiency. In order to calculate scale inefficiency, the fraction $\eta^*_{CCR}/\eta^*_{BCC}$ should be calculated.

Cooper, Lawrence, & Tone (2006) indicates the following formula in their book as CCR output-oriented model.

$$\text{Max } \eta_{CCR} \tag{4.5}$$

$$\text{Subject to } \sum_{i=1}^{m} x_{i0} - \sum_{i=1}^{m} \lambda_j \, x_{ij} \geq 0 \;\; \forall j \tag{4.6}$$

$$\eta_{CCR} \sum_{r=1}^{s} y_{r0} - \sum_{r=1}^{s} \lambda_j \, y_{rj} \leq 0 \qquad \forall j \tag{4.7}$$

$$\lambda_j \geq 0$$

where

CCR : CCR model

i Index number of inputs i=1, 2,..,m

r Index number of outputs r=1,2,…,s

j Index number of decision units j=1,2,..,N

η_{CCR} = Efficiency Score

y_{rj} = Quantity of output r for the decision unit j

y_{r0} = Quantity of output r for the evaluated decision unit

x_{ij} = Quantity of input i for the decision unit j

x_{i0} = Quantity of input i for the evaluated decision unit

λ_j = The benchmark index for the decision unit j

Equation (4.5) is the objective function to maximize η_{CCR} for increasing y_{r0} to $\eta_{CCR}y_{r0}$ without increasing x_{i0}. Equation (4.6) and Equation (4.7) confirm that the improved decision unit is located in the feasible region. The evaluated decision unit with the benchmark index equal to 1 and consequently η_{CCR} equal to 1 is efficient. The evaluated decision unit with η_{CCR} greater than 1 is inefficient and the benchmarks for this inefficient decision unit are the decision units with λ_j greater than zero.

4.4.1.3. Mix inefficiency

One of the properties of DEA is its capability to calculate the modification that should be done in the inputs and outputs of the inefficient decision units to make them efficient. Cooper, Lawrence, & Tone (2006) indicates the following two-phase formula in their book to calculate mix-inefficiency.

Phase one:

$$\text{Max } \eta_{CCR} \tag{4.8}$$

$$\text{Subject to } \sum_{i=1}^{m} x_{i0} - \sum_{i=1}^{m} \lambda_j x_{ij} - \sum_{i=1}^{m} s_i^- = 0 \quad \forall j \tag{4.9}$$

$$\eta_{CCR} \sum_{r=1}^{s} y_{r0} - \sum_{r=1}^{s} \lambda_j y_{rj} + \sum_{r=1}^{s} s_r^{+} = 0 \qquad \forall j \qquad (4.10)$$

$$\lambda_j \geq 0$$

where

i	Index number of inputs	i=1, 2,..,m
r	Index number of outputs	r=1,2,…,s
j	Index number of decision units	j=1,2,..,N

η_{CCR} = Efficiency Score

y_{rj} = Quantity of output r for the decision unit j

y_{r0} = Quantity sdof output r for the evaluated decision unit

x_{ij} = Quantity of input i for the decision unit j

x_{i0} = Quantity of input i for the evaluated decision unit

λ_j = The benchmark index for the decision unit j

s_i^{-}= Excess in input i

s_r^{+}= Shortfall in output r

Equations (4.8)-(4.11) are the same as Equation (4.1)-(4.4). The only difference is that in Equation (4.9) and Equation (4.10) the inequality constraints are changed to quality constraints by adding variables for input excess and output shortfall. . In phase one, $\eta_{CCR}$is calculated to be substitud in phase two for obtaining mix inefficiency which is the sume of S^{-*} and S^{+*}.

Phase two:

$$\text{Max } w = \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \tag{4.12}$$

$$\text{Subject to } \sum_{i=1}^{m} s_i^- = x_0 - \sum_{i=1}^{m} \lambda_j x_{ij} \qquad \forall j \tag{4.13}$$

$$\sum_{r=1}^{s} s_r^+ = \sum_{r=1}^{s} \lambda_j y_{rj} - \eta_{CCR}^{*} y_0 \qquad \forall j \tag{4.14}$$

$$\lambda_j \geq 0, \quad s_i^- \geq 0, s_r^+ \geq 0$$

where

w = Mix inefficiency

η_{CCR}^{*} = Maximum technical Efficiency Score from Phase one

Equation (4.12) is the objective function to maximize the sum of input excess and output shortfall which is mix inefficiency. Equation (4.13) is the constraint for input excess and Equation (4.14) is the constraint for output shortfall. If $w^* \neq 0$, it means $S^{-*} \neq 0$ or (and) $S^{+*} \neq 0$ that shows there is mix inefficiency.

4.4.2. Identify efficient configurations

The configuration is identified as efficient if the configuration does not have technical inefficiency, scale inefficiency and mix inefficiency.

In this section, efficient configurations are discriminated from inefficient configurations through identifying the terms of inefficiency.

4.4.3. Selecting the most efficient configuration

It is highly possible that the applied DEA model in section 4.4.2 identifies multiple configurations as efficient. Unrealistic weight distribution is one of the reasons for the presence of multiple efficient configurations (Li & Reeves, 1999). Unrealistic weight distribution is one of the drawbacks of DEA and is defined as assigning an unrealistic too large weight to a single outputs or/and assigning a too small weight to single input to make the configuration relatively efficient (Li & Reeves, 1999). Li & Reeves (1999), proposed a minimax approach in their study to overcome the problem of unrealistic weight distirbution which is formulated as follow:

$$\text{Min} \quad M \tag{4.14}$$

$$\text{Subject to} \quad \sum_r w_i x_{i0} = 1 \tag{4.15}$$

$$\sum_r u_r y_{rj} - \sum_r w_i x_{ij} + d_j = 0 \qquad \forall j \tag{4.16}$$

$$M - d_j \geq 0 \qquad \forall j \tag{4.17}$$

$$u_r, w_i, d_j \geq \varepsilon \geq 0 \qquad r = 1,\ldots,s;\ i=1,\ldots,m;\ j = 1,\ldots,n$$

Equation (4.14) minimizes M which is the maximum of all deviation variables. Equation (4.15) is the constraint to linearize the fractional objective function of basic DEA. Equation (4.16) makes the efficiency score be less than one and d_j is the deviation variable for each configuration. Equation (4.17) makes M be the maximum of all the deviations. The efficiency score is calculated as h_0 = 1-d_0 and the configuration is minimax efficient if and only if the value d_0 that minimizes M is zero (Ertay, Ruan, & Tuzkaya, 2006).

Minimax efficiency approach discriminates efficient configurations from inefficient configurations more realistically compare to the classic DEA approach (the approach in Section 4.4.2). If the origin of inefficiency is not required to be identified, minimax efficiency approach is only applied to select efficient configurations.

The selected efficient configurations through both minimax efficiency approach and classic DEA approach should have the deviation variable equal to zero. The minimax efficient configurations have an additional constraint which is the minimizing the maximum of the all the deviation variables of other configurations should be minimized.

The presence of the additional constraint makes the selected minimax efficient configurations more realistic.

The objective function of minimax efficiency approach (Equation (4.14)) can be modified to "M - kd_0 ". Firstly, the minimax efficiency approach is applied with the coefficient k equal to zero. If more than one configuration is selected as efficient, k ranges between (0, 1) and determined by trial and error. The number of trials to find the most efficient configuration is limited and the first trial that can provide only one configuration as the most efficient is selected as the answer. The modified minimax efficiency score is less than the original minimax efficiency score so the modified minimax efficiency approach can converge multiple efficient configurations to one efficient configuration (Ertay, Ruan, & Tuzkaya, 2006).

The reason for selecting only one configuration as the most efficient is to provide a benchmark. Benchmark configuration is defined as the superior configuration that provide robustness to demand variability and the balnce between indicators of efficiency. In case, the hospital is not able to change its current configuration, presence of the benchmark leads to continuous improvement of the implemented configuration (Lai, Huang, & Wang, 2010).

CHAPTER 5. CASE STUDY

In this chapter, we conduct an experiment and test different settings in terms of multiple numbers of appointments to explain how the proposed methodology can be used as well as the applicability of the demonstrated methodology.

5.1. Clinic settings

The setting in this study includes two primary cares, three specialty clinics and one laboratory. Demand pattern is defined in terms of arrival rate of requested appointment, appointments' type and percentage of multiple appointments versus single appointment for the combination of clinics.

We defined three types of appointment, which are patient appointment calls, walk-in patients and provider requested appointments. In this study, patient appointment calls and provider requested appointment are considered. Walk-in appointments involve considerable stochastic behaviors and bring significant variability to the system that would be studied in future. It is assumed that provider appointment request is transferred to the related scheduler through phone call.

The mean time between arrivals of the merged patient appointment calls and provider appointment requests is 2 minutes and follows exponential distribution. Service time follows triangular distribution. The distributions for service time and arrival of appointment request are obtained from analyzing the historical data.

The clinics' interaction index as defined earlier explains the impact of clinic setting complexity on scheduling models. Multiple appointments' intensity is one of the major factors in clinic setting complexity. In this study clinic setting complexity is limited to multiple appointments' intensity. In future, we consider clinic setting complexity as multi dimensional array that would consider other factors in addition to multiple appointments' intensity.

Clinics' interaction index is calculated from the ratio of total number of multiple appointments over total number of appointments. In this study, three clinic settings with interaction index of 91 percent, 40 percent and 10 percent is studied. The first clinic setting has high, the second one has medium and the last one has low interaction level.

Tables 5.1, 5.2 and 5.3 show the percentage of different multiple appointments for the three settings in this study.

Table 5.1. Interaction level for the setting with high interaction.

Number of appointments	**Frequency**
5	14%
4	20%
3	27%
2	30%
1	8%

Table 5.2. Interaction level for the setting with medium interaction.

Number of appointments	**Frequency**
5	3%
4	1%
3	1%
2	35%
1	60%

Table 5.3. Interaction level for the setting with low interaction.

Number of appointments	Frequency
5	0%
4	0%
3	5%
2	5%
1	90%

5.2. Configuration generation

Different configurations of the three major appointment scheduling models CSM, DSM and HSM are generated for different type of clinic settings, which are considered as decision units for DEA models.

The generated configurations for efficiency evaluating in this study are: two centralized scheduling models (CSM). The first CSM includes a call center and the schedulers are single task, the second CSM, does not include a call center and the schedulers are multitasking. As a result for the second CSM, the utilization of the schedulers should be multiplied by a coefficient to obtain the correct value for utilization of staffs for performing scheduling task. The third and forth configurations are DSM. Six configurations for HSM are also developed that assign clinics to different clusters and each cluster has its own scheduler.

The configurations descriptions for the three types of settings are illustrated in Table 5.4 and Table 5.5.

5.3. Simulation study

Simulation study is conducted to obtain the outputs' value for the applied Data Envelopment Analysis models in this study. Simulation models for the configurations of clinic settings are created by the simulation package arena (Kelton, Sadowski, & Swets, 2010).

Table 5.4. Configuration generation for the major scheduling models.

Decision Units	Description
DU1	CSM (1) with call center and single task schedulers
DU2	CSM (2) with call center and multitasking schedulers
DU3	DSM (1) with single task schedulers
DU4	DSM (2) with multitasking schedulers
DU5-DU10	HSM

Three different clinic settings are studied and nine configurations for the scheduling models are generated for each clinic setting. As a result, twenty seven simulation models with thirty replications are designed. The simulation model is run for 8 hours. The possible number of multiple appointments is 5, 4, 3 and 2 appointments.

5.3.1. Simulation outputs

The simulation output values for the setting with high interaction level, medium interaction level and low interaction level are illustrated in Table 5.6 Table 5.7 and Table 5.8.

5.3.2.Validation/Verification

Model verification is applied to compare the conceptual model with the simulated model. The selected verification method in this study is sensitivity analysis.

Sensitivity analysis is performed to see the behavior of some performance measures by changing firstly the inter-arrival between appointment requests and secondly proportion of multiple appointments.

The inter-arrival time is changed from 1 to 3.5 and the behavior of average waiting time before connecting to a scheduler is captured. Figure 5.1 shows the obtained waiting time by the simulation model. As it is observed, the waiting time is decreased by increasing the inter-arrival time which match the expectation.

Table 5.5. Configurations with hybrid scheduling model.

Decision Units	HSM	Cluster1	Cluster2	Cluster3
DU5	HSM (1)	Lab, primary one, urology	Lab, primary two, physical therapy	Lab, Audiology
DU6	HSM (2)	Lab, primary one, primary two	Urology, audiology	Physical therapy
DU7	HSM (3)	Lab, primary one, primary two	Lab, urology, physical therapy	Audiology
DU8	HSM (4)	Lab, primary one, primary two, urology	Audiology, Physical Therapy	
DU9	HSM (5)	Lab, primary one, urology, physical Therapy	Lab, primary2, audiology	
DU10	HSM (6)	Lab, primary one, Audiology, urology	Lab, primary2,Physical therapy	

The inter-arrival time is changed from 1 to 3.5 and the behavior of holding time to schedule all the requested appointments is captured. Figure 5.2 shows the obtained holding time by the simulation model. As it is observed, holding time are decreased by increasing the inter-arrival time which match the expectation.

Table 5.6. Simulation outputs for the setting with high interaction.

Decision Unit	Model	Avg WT	Avg Total time	Avg Utilization rate	Number of Staffs
DU1	CSM (1)	0.00925	0.05119	0.42	3
DU2	CSM (2)	0.00925	0.05119	0.6	3
DU3	DSM (1)	0.0549	0.1424	0.43	5
DU4	DSM (2)	0.0549	0.1424	0.65	5
DU5	HSM (1)	0.1111	0.1723	0.62	3
DU6	HSM (2)	0.0867	0.1506	0.53	3
DU7	HSM (3)	0.0777	0.1424	0.51	3
DU8	HSM (4)	0.02823	0.0903	0.47	2
DU9	HSM (5)	0.03184	0.0874	0.48	2
DU10	HSM (6)	0.0414	0.0999	0.49	2

Table 5.7. Simulation outputs for the setting with medium interaction.

Decision Unit	Model	Avg WT	Avg Total time	Avg Utilization rate	Number of Staffs
DU1	CSM (1)	0.003032	0.0320	0.41	3
DU2	CSM (2)	0.003032	0.0320	0.6	3
DU3	DSM (1)	0.01268	0.0545	0.2	5
DU4	DSM (2)	0.01268	0.0545	0.4	5
DU5	HSM (1)	0.01882	0.05236	0.33	3
DU6	HSM (2)	0.016034	0.04856	0.256	3
DU7	HSM (3)	0.0186	0.05390	0.3	3
DU8	HSM (4)	0.03066	0.06604	0.419	2
DU9	HSM (5)	0.02368	0.05525	0.387	2
DU10	HSM (6)	0.00599	0.03573	0.3	2

Table 5.8. Simulation outputs for the setting with low interaction.

Decision Unit	Model	Avg WT	Avg Total time	Avg Utilization rate	Number of Staffs
DU1	CSM (1)	0.002372	0.0288	0.3	3
DU2	CSM (2)	0.002372	0.0288	0.5	3
DU3	DSM (1)	0.006516	0.036227	0.35	5
DU4	DSM (2)	0.006516	0.036227	0.65	5
DU5	HSM (1)	0.014799	0.0421904	0.25	3
DU6	HSM (2)	0.011171	0.038508	0.21	3
DU7	HSM (3)	0.0172	0.045078	0.21	3
DU8	HSM (4)	0.0171	0.04487	0.2	2
DU9	HSM (5)	0.01942	0.04619	0.2	2
DU10	HSM (6)	0.0033	0.028	0.21	2

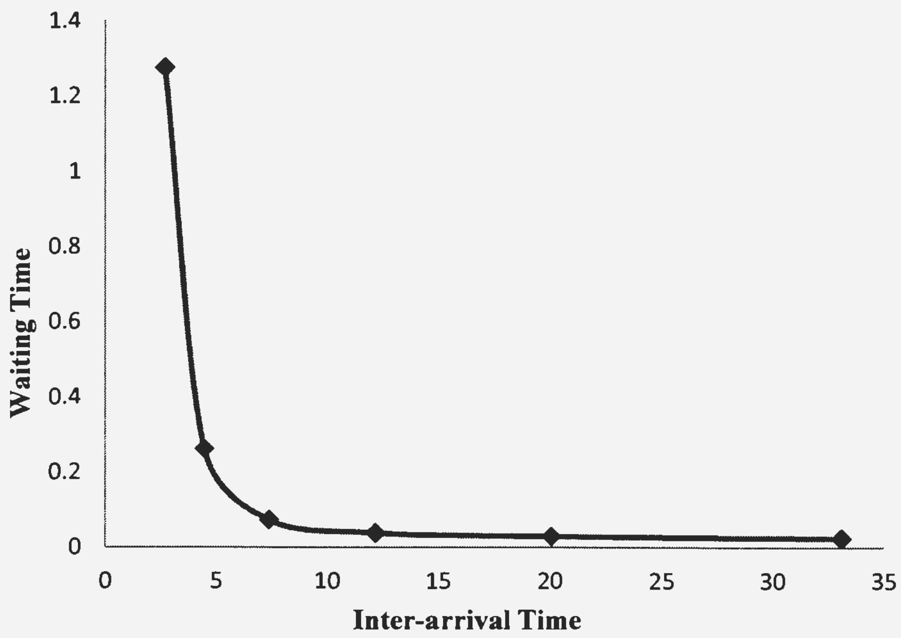

Figure 5.1. The impact of request arrival on waiting time.

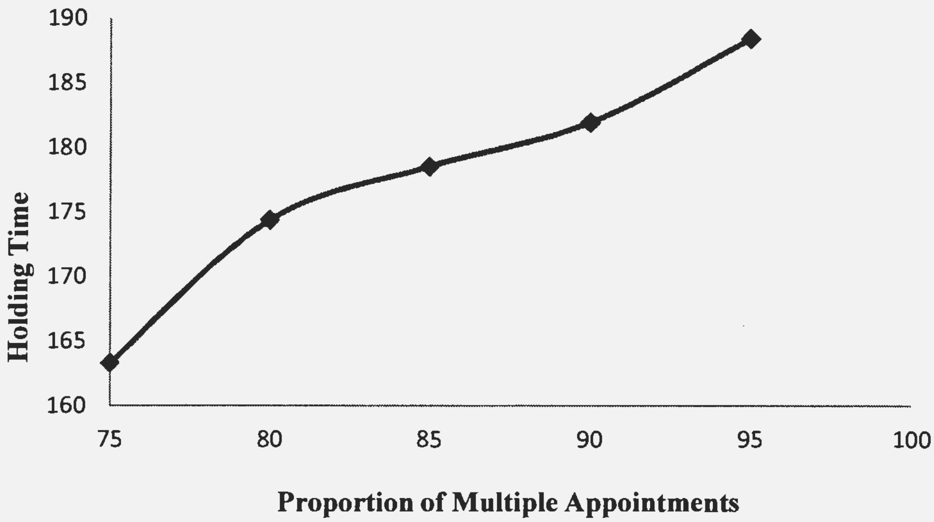

Figure 5.2. The impact of request arrival on holding time.

The proportion of multiple appointments is changed from 75 to 95 percent and the behavior of average waiting time before connecting to a scheduler as well as average holding time are captured. Figure 5.3 and Figure 5.4 show the obtained waiting time and holding time by the simulation model. As it is observed, both the waiting time and holding time are increased by increasing the inter-arrival time which match the expectation.

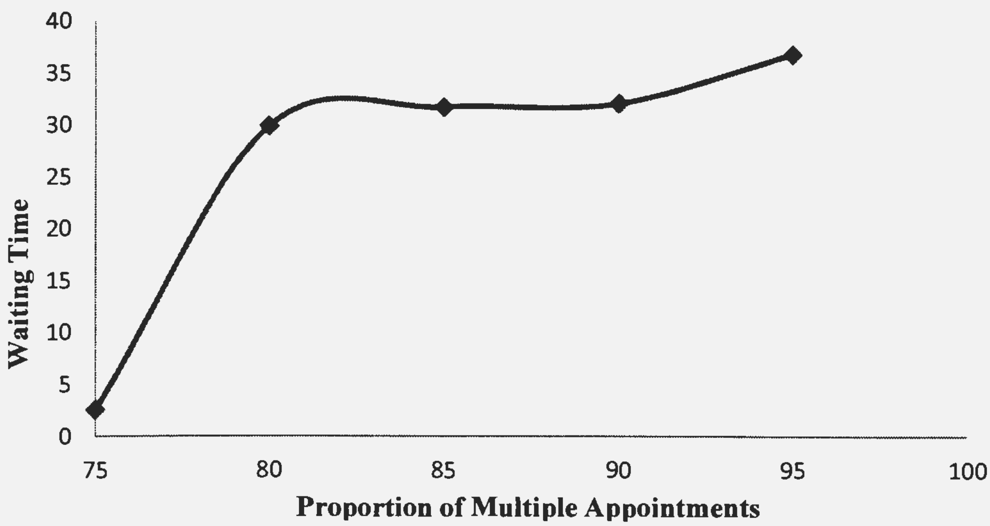

Figure 5.3. The impact of proportion of multiple appointments on waiting time.

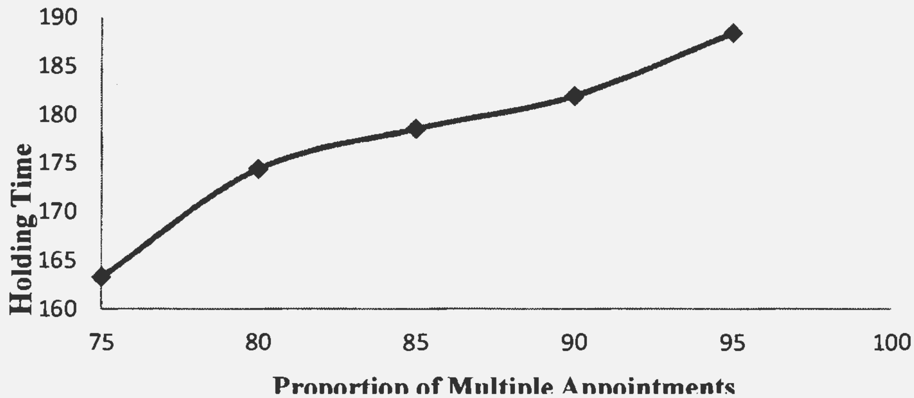

Figure 5.4. The impact of proportion of multiple appointments on holding time.

5.4. Results

5.4.1. Analyze inefficient configurations

5.4.1.1. Clinic setting with high interaction level

The DEA inputs and outputs value for the setting with high interaction level are illustrated in Table 5.9 and Table 5.10.

Table 5.9. DEA inputs for the setting with high interaction.

Decision Unit	Model	Number of staffs	Training cost	Software cost
DU1	CSM (1)	3	400 * 3 = 1200	900
DU2	CSM (2)	3	450*3 = 1350	900
DU3	DSM (1)	5	150 * 5 = 750	700
DU4	DSM (2)	5	200*5=1000	700
DU5	HSM (1)	3	350*2 + 200 = 900	800
DU6	HSM (2)	3	350 + 2*200 = 750	800
DU7	HSM (3)	3	2*350 + 200 = 900	800
DU8	HSM (4)	2	375 + 100 = 475	850
DU9	HSM (5)	2	375 + 200 = 575	850
DU10	HSM (6)	2	375 + 200 = 575	800

Through analyzing different aspects of efficiency for the 10 generated configuration of the clinic setting with high interaction level. Equations (4.1)-(4.4) are applied to evaluate technical efficiency. Configurations 5 and 7 are technically inefficient that indicates the scheduling software and the schedulers do not use their available capability.

Equations (4.5)-(4.7) are applied to evaluate scale efficiency. Scale inefficiency in the configurations 3, 5, 6, 7 and 9 shows the complexity of scheduling software and schedulers' skill level in these two configurations are not sufficient and improvement should be applied in the scheduling software and schedulers' skills.

Table 5.10. DEA outputs for the setting with high interaction.

Decision Unit	Model	Avg WT Reciprocal	Avg Total time Reciprocal	Avg Utilization
DU1	CSM (1)	108.12	19.54	0.42
DU2	CSM (2)	108.12	19.54	0.6
DU3	DSM (1)	18.22	7.02	0.43
DU4	DSM (2)	18.22	7.02	0.65
DU5	HSM (1)	9	5.8	0.62
DU6	HSM (2)	11.53	6.64	0.53
DU7	HSM (3)	12.87	7.02	0.51
DU8	HSM (4)	35.42	11.07	0.47
DU9	HSM (5)	31.41	11.44	0.48
DU10	HSM (6)	24.16	10.01	0.49

Finally, Equations (4.8)-(4.14) are applied to evaluate mix inefficiency. Mix inefficiency in configuration 3, 5, 6, 7 and 9 is an evidence for the presence of extra inputs and outputs' shortfalls. Table 5.11 presents the results from analyzing different terms of inefficiency for the clinic setting with high interaction level.

5.4.1.2. Clinic setting with medium interaction level

The DEA inputs and outputs value for the setting with medium interaction level are illustrated in Table 5.12 and Table 5.13.

Through analyzing different aspects of efficiency for the 10 generated configurations of the clinic setting with medium interaction level. Equations (4.1)-(4.4) are applied to evaluate technical efficiency. Configurations 5 and 7 are technically inefficient that indicates the scheduling software and the schedulers do not use their capability.

Table 5.11. Inefficiency terms for the setting with high interaction.

Decision Unit	Model	Technical efficiency score	Scale efficiency score	Mix efficiency score
DU1	CSM (1)	1*	1*	0*
DU2	CSM (2)	1*	1*	0*
DU3	DSM (1)	1*	1.37	18.1
DU4	DSM (2)	1*	1*	0*
DU5	HSM (1)	1.03	1.03	71.72
DU6	HSM (2)	1*	1.11	36
DU7	HSM (3)	1.28	1.03	49.07
DU8	HSM (4)	1*	1*	0*
DU9	HSM (5)	1*	1.09	73.6
DU10	HSM (6)	1*	1*	0*

Equations (4.5)-(4.7) are applied to evaluate scale efficiency. Scale inefficiency in the configurations 3, 4, 5, 6, 7 and 9 shows the complexity of scheduling software and schedulers' skill level are not sufficient and improvement should be applied in the scheduling software and schedulers' skills.

Finally, Equations (4.8)-(4.14) are applied to evaluate mix inefficiency. Mix inefficiency in configurations 3, 4, 5, 6, 7 and 9 is an evidence for the presence of extra inputs and output shortfalls. Table 5.14 presents the results from analyzing different terms of inefficiency for the clinic setting with high interaction level.

Table 5.12. DEA inputs for the setting with medium interaction.

Decision Unit	Model	Number of staffs	Training cost	Software cost
DU1	CSM (1)	3	400 * 3 = 1200	900
DU2	CSM (2)	3	450*3=1350	900
DU3	DSM (1)	5	150 * 5 = 750	700
DU4	DSM (2)	5	200*5=1000	700
DU5	HSM (1)	3	350*2 + 200 = 900	800
DU6	HSM (2)	3	350 + 2*200 = 750	800
DU7	HSM (3)	3	2*350 + 200 = 900	800
DU8	HSM (4)	2	375 + 100 = 475	850
DU9	HSM (5)	2	375 + 200 = 575	850
DU10	HSM (6)	2	375 + 200 = 575	800

5.4.1.3. Clinic setting with low interaction level

The DEA inputs and outputs value for the setting with low interaction level are illustrated in Table 5.15 and Table 5.16.

Through analyzing different aspects of efficiency for the 10 generated configuration of the clinic setting with low interaction level. Equations (4.1)-(4.4) are applied to evaluate technical efficiency. 10 runs of the model should be performed to identify technical efficient configurations. Configurations 1, 5, 6, 7 and 9 are technically inefficient which is proof for inefficiency in resource allocation in this configuration.

Equations (4.5)-(4.7) are applied to evaluate scale efficiency. 10 runs and 10 comparisons of the model should be performed to identify scale efficient configuration. Scale inefficiency in the configurations 1, 5, and 9 shows the complexity of scheduling software and schedulers' skill level in these two configurations are not sufficient and improvement should be applied in the scheduling software and schedulers' skills.

Table 5.13. DEA outputs for the setting with medium interaction.

Decision Unit	**Model**	**Avg WT Reciprocal**	**Avg Total time Reciprocal**	**Avg Utilization**
DU1	CSM (1)	333.33	31.25	0.41
DU2	CSM (2)	333.33	31.25	0.6
DU3	DSM (1)	78.86	18.25	0.2
DU4	DSM (2)	78.86	18.25	0.4
DU5	HSM (1)	53.13	19.1	0.33
DU6	HSM (2)	62.37	20.59	0.256
DU7	HSM (3)	53.76	18.55	0.3
DU8	HSM (4)	32.62	15.14	0.419
DU9	HSM (5)	42.23	18.01	0.387
DU10	HSM (6)	166.95	27.99	0.3

Finally, Equations (4.8)-(4.14) are applied to evaluate mix inefficiency. The mix inefficient configurations are identified through 20 runs of model. Mix inefficiency in configurations 5, 6, 7 and 9 is an evidence for the presence of extra inputs and outputs' shortfalls. Table 5.17 presents the results from analyzing different terms of inefficiency for the clinic setting with high interaction level.

Table 5.14. Inefficiency terms for the setting with medium interaction.

Decision Unit	Model	Technical efficiency score	Scale efficiency score	Mix efficiency score
DU1	CSM(1)	**1***	**1***	**0***
DU2	CSM(2)	**1***	**1***	**0***
DU3	DSM(1)	**1***	1.3	132.97
DU4	DSM(2)	**1***	1.33	232.76
DU5	HSM (1)	1.35	1.06	179.56
DU6	HSM (2)	**1***	1.15	174.26
DU7	HSM (3)	1.39	1.06	188.57
DU8	HSM (4)	**1***	**1***	**0***
DU9	HSM (5)	**1***	1.09	143
DU10	HSM (6)	**1***	**1***	**0***

Table 5.15. DEA inputs for the setting with low interaction.

Decision Unit	Model	Number of staffs	Training cost	Software cost
DU1	CSM (1)	3	400 * 3 = 1200	900
DU2	CSM (2)	3	450*3 = 1350	900
DU3	DSM(1)	5	150 * 5 = 750	700
DU4	DSM(2)	5	200*5=1000	700
DU5	HB(1)	3	350*2 + 200 = 900	800
DU6	HB(2)	3	350 + 2*200 = 750	800
DU7	HB(3)	3	2*350 + 200 = 900	800
DU8	HB(4)	2	375 + 100 = 475	850
DU9	HB(5)	2	375 + 200 = 575	850
DU10	HB(6)	2	375 + 200 = 575	800

Table 5.16. DEA outputs for the setting with low interaction level.

Decision Unit	Model	Avg WT Reciprocal	Avg Total time Reciprocal	Avg Utilization
DU1	CSM(1)	434.79	35.71	0.3
DU2	CSM(2)	434.79	35.71	0.5
DU3	DSM(1)	192.31	31.25	0.35
DU4	DSM(2)	192.31	31.25	0.65
DU5	HB(1)	66.67	25	0.25
DU6	HB(2)	133.33	29.94	0.21
DU7	HB(3)	188.68	32.26	0.21
DU8	HB(4)	454.54	35.71	0.2
DU9	HB(5)	500	37.04	0.2
DU10	HB(6)	303.03	35.7	0.21

Table 5.17. Inefficiency terms for the setting with low interaction.

Decision Unit	Model	Technical efficiency score	Scale efficiency score	Mix efficiency score
DU1	CSM (1)	1.02	1.08	**0***
DU2	CSM (2)	**1***	**1***	**0***
DU3	DSM (1)	**1***	**1***	**0***
DU4	DSM (2)	**1***	**1***	**0***
DU5	HSM (1)	1.4	1.01	324.7
DU6	HSM (2)	1.04	**1***	58.7
DU7	HSM (3)	1.1	**1***	385.67
DU8	HSM (4)	**1***	**1***	**0***
DU9	HSM (5)	1.07	1.06	142
DU10	HSM (6)	**1***	**1***	**0***

5.4.2. Identify efficient configurations

5.4.2.1. Clinic setting with high interaction level

The configuration is identified as efficient if it is efficient in terms of technical, scale and mix. 20 comparisons are done between terms of efficiency and the identified efficient configurations for clinic setting with high interaction level are: DU1, DU2, DU4, DU8, and DU10.

5.4.2.2. Clinic setting with medium interaction level

The efficient configurations for clinic setting with medium interaction level are: DU1, DU2, DU8, and DU10.

5.4.2.3. Clinic setting with low interaction level

The efficient configurations for clinic setting with low interaction level are: DU1, DU2, DU3, DU4, DU8, and DU10.

5.4.3. Selecting the most efficient configuration

5.4.3.1. Clinic settings with high interaction level

Equation (4.14)-(4.17) are applied to remove unrealistic efficient configurations through minimax efficiency approach. Table 5.18 presents the results for selecting the most efficient configuration in the clinic setting with high interaction level. The DEA model runs 10 times and configuration 2 is identified as the most efficient configuration which indicates in the presence of high proportion of multiple appointments advanced information sharing system and multi-tasking schedulers with high skill level are required.

The properties of the configuration that is selected as the most efficient one for the setting with high interaction level are presented in Table 5.19. Single tasking versus multitasking is one of the major differences between configurations 1 and 2. Configuration 2 with multitasking option is applicable if the schedulers' utilization in configuration 1 is less than 50 percent. In a case the utilization is smaller than 50 percent, a coefficient should be multiplied to the schedulers' utilization in the first configuration to obtain the second configuration's utilization. In the second configuration 6 staffs do the task of 9 staffs (3 schedulers and 6 check-in staffs), so there is 0.33 percent ((6-9)/9) decrease in staffs number that makes staffs' utilization (42 percent) increase to 60 percent.

Table 5.18. Selecting the most efficient configuration for the high interaction setting.

Decision Unit	**Model**	**Minimax efficiency score**
DU1	CSM (1)	0.75
DU2	CSM (2)	**1***
DU3	DSM (1)	0.65
DU4	DSM (2)	0.9
DU5	HSM (1)	0.78
DU6	HSM (2)	0.83
DU7	HSM (3)	0.69
DU8	HSM (4)	0.62
DU9	HSM (5)	0.73
DU10	HSM (6)	0.72

5.4.3.2. Clinic settings with medium interaction level

Equation (4.14)-(4.17) are applied to remove unrealistic efficient configurations. The minimax efficiency approach selects the two Configurations 2 and 10 as efficient configurations so the modified minimax efficiency approach with k equal to 0.25 is applied to determine the most efficient configuration. Configuration 10 is selected as the most efficient configuration.

Configuration 10 with HSM decision structure outweighs configuration 2 with CSM decision structure in terms of lower cost. In the presence medium proportion of multiple appointments in the current clinic setting, it is efficient to form two clusters and assign the clinics with high interaction level to the same clinic with CSM decision structure. Configuration 10 provides sufficient accessibility and resource utilization besides lower cost. Table 5.20 presents the results for selecting the most efficient configuration in the clinic setting with medium interaction level.

The properties of the configuration that is selected as the most efficient one for the setting with medium interaction level are presented in Table 5.21.

5.4.3.3. Clinic settings with low interaction level

Equations (4.14)-(4.17) are applied to select the most efficient configuration through minimax efficiency approach. Configurations 4 and 10 are selected so the modified minimax efficiency approach with k equal to 0.25 is applied and configuration 4 is selected as the most efficient configuration. Configuration 4 has DSM decision structure outweighs configuration 10 with HSM decision structure in terms of lower cost. Table 5.22 presents the results for selecting the most efficient configuration in the clinic setting with low interaction level.

The properties of the configuration that is selected as the most efficient one for the setting with low interaction level are presented in Table 5.23.

Table 5.19. Properties of the final configuration for the setting with high interaction.

Scheduling Model	Configuration Type	Avg WT	Avg total time	Avg Utilization
CSM (2)	CSM with multitasking staffs and without scheduling center	33.3 sec	3.07 min	0.6

Table 5.20. Selecting the most efficient configuration for the setting with medium interaction.

Decision Unit	Model	MiniMax efficiency score	efficiency score of M – 0.25*d0
DU1	CSM (1)	0.89	0.72
DU2	CSM (2)	**1***	0.864
DU3	DSM (1)	0.628	0.628
DU4	DSM (2)	0.64	0.65
DU5	HB (1)	0.678	0.678
DU6	HB (2)	0.823	0.789
DU7	HB (3)	0.65	0.65
DU8	HB (4)	0.62	0.62
DU9	HB (5)	0.69	0.69
DU10	HB (6)	**1***	**1***

Table 5.21. Properties of the final configuration for the setting with medium interaction.

Scheduling Model	Configuration Type	Avg WT	Avg total time	Avg Utilization
HSM (6)	HSM with 2 cluster	21.56 sec	2.14 min	0.3

Table 5.22. Selecting the most efficient configuration for the setting with low interaction.

Decision Unit	Model	MiniMax efficiency score	efficiency score of M – 0.25*d0
DU1	CSM (1)	0.95	0.75
DU2	CSM (2)	0.89	0.78
DU3	DSM (1)	0.9	0.7
DU4	DSM (2)	**1***	**1***
DU5	HSM (1)	0.88	0.7
DU6	HSM (2)	0.87	0.94
DU7	HSM (3)	0.73	0.73
DU8	HSM (4)	0.74	0.74
DU9	HSM (5)	0.71	0.71
DU10	HSM (6)	**1***	0.99

Table 5.23. Properties of the final configuration for the setting with low interaction.

Scheduling Model	Configuration Type	Avg WT	Avg total time	Avg Utilization
DSM (2)	Multitasking staffs	23.45 sec	2.17 min	0.65

CHAPTER 6. CONCLUSION AND FUTURE RESEARCH

This study is developed to select the most efficient scheduling model for a specific clinic setting to address the significance of clinic setting context on the efficiency of appointment schedùling models. The clinic setting with high proportion of single appointments and CSM scheduling model, impose a unnecessary high cost to the clinic because the clinics does not need sharing information system and coordination. The clinic setting with high proportion of multiple appointments and DSM scheduling model causes huge patient dissatisfaction because the clinic setting would take a long time to obtain the required information and coordinate between different clinics.

There are three major patient appointment scheduling models: centralized scheduling model (CSM), decentralized scheduling model (DSM) and hybrid scheduling model (HSM). The decision structure for CSM is that the requested multiple appointments are handled only by one scheduler and there is not any difference between different schedulers.

The decision structure for DSM is that the incoming requests should be directed to the schedulers of the specific clinics. The requested multiple appointments are handled by the schedulers of the related clinics so multiple telephone contacts are required.

The decision structure for hybrid scheduling model (HSM) is the requested appointments in some clinics are scheduled based on CSM and in some other clinic are scheduled based on DSM. Different clusters are formed and each cluster includes one or multiple number of clinics. The clinics in the same cluster have CSM decision structure, but different clusters act like individual clinics with DSM decision structure.

In order to select the most efficient configuration for providing an efficient appointment scheduling system, a procedure is developed in this study based on the integration of simulation and data envelopment analysis (DEA).

The reason for conducting simulation is that different configurations cannot be applied for the specific clinic setting. In order to obtain the required outputs, different configurations of scheduling models are simulated. The outputs are selected based on the categories of efficiency indicators for appointment scheduling model.

There are three categories of efficiency indicators for appointment scheduling models: stakeholders' satisfaction, resource utilization, and cost. The selected clinic setting's outputs in this study are patient satisfaction, schedulers' utilization, scheduling software cost and schedulers' training cost.

DEA approach is applied to compare different configurations of scheduling models for a specific clinic setting and select the most efficient configurations. There are three major elements for DEA: Decision Units (DU), DEA inputs and DEA outputs. The units that are compared through DEA approach are Decision Units that are different models for appointment scheduling models in this study.

DEA inputs are the indicators that should be minimized like cost in this study. DEA outputs are the indicators that should be maximized that are resource utilizations and patient satisfaction in this study.

Patient satisfaction is a qualitative indicator and is determined by measuring accessibility. Accessibility is estimated through waiting time before connecting to a scheduler and call duration to schedule the requested appointments. In order to assign accessibility as DEA output, the reciprocal of time should be considered.

The first stage for applying the DEA approach is selecting the proper model. There are two major DEA models: Charnes-Cooper-Rhodes (CCR) and Banker-Charnes-Cooper (BCC).

In order to determine the most efficient configuration, three terms of efficiency should be calculated. The efficiency terms are: technical efficiency, scale efficiency and mix efficiency.

Technical efficiency is calculated through BCC output-oriented model and presents scheduling software and schedulers' performances match their capability. Scale efficiency is calculated through both CCR output-oriented model and BCC output-oriented model and shows complexity of the scheduling software and skill level of schedulers match the clinic setting context. The mix efficiency is calculated through the two phase CCR output-oriented model and indicates the presence of extra inputs and outputs' shortfalls.

The configuration is efficient if it has all the three terms of efficiency. If only one configuration is identified as efficient, this configuration is also the best configuration. But if more than one configuration is identified as efficient, there is a possibility that DEA identified some configuration as efficient unrealistically. In order to remove unrealistic efficient configurations, minimax efficiency approach is applied. If the minimax efficiency approach identified more than one configuration as efficient, the modified minimax efficiency approach is applied to select only one configuration as the most efficient configuration.

In order to evaluate the impact of arrival rate of requested appointments and proportion of multiple appointments on the efficiency of appointment scheduling system, the interaction level index is defined. The interaction level index is the ratio of proportion of multiple appointments over total number of requested appointments. In this study, interaction level index differentiate between clinic settings. We defined three different clinic settings with low interaction level (less than 15%), medium interaction level (from 25% to 50%) and high interaction level (more than 70%).

A case study is conducted in this research as a guide to use and as a proof for the validity of the developed procedure. Three clinic settings are studied: the clinic setting with low interaction level (5%), the clinic setting with medium interaction level (40%) and the clinic setting with high interaction level (92%).

For the clinic setting with high interaction level in this study, the second configuration of centralized scheduling model (the schedulers are multitasking and are distributed between individual clinics) is selected as the most efficient scheduling model. For the clinic setting with medium interaction level in this study, the sixth configuration of hybrid scheduling model (the clinics are assigned into two clusters) is selected as the most efficient scheduling model. For the clinic setting with low interaction level in this study, the second configuration of decentralized scheduling model (the schedulers are multitasking) is selected as the most efficient configuration. The most efficient scheduling models for the three clinic settings in this study are selected in terms of the balance between patient satisfaction, schedulers' utilization and scheduling system cost.

The proposed methodology could select the most efficient configuration for three different clinic setting. The selected clinic settings match the expectation which validates the proposed procedure for selecting the most efficient scheduling model configuration for different types of clinic settings.

For the future work, further simulation study would be done to determine the mutual impact of scheduling model and clinic flows and also additional inputs and outputs would be defined to apply in the DEA approach and select the most efficient patient appointment scheduling model.

REFERENCES

1. Agnihotri, S., & Taylor, P. (1991). Staffing a centralized Appointment Scheduling Department in Lourdes Hospital. Interfaces , 1-11.
2. Avkiran, N. (2001). Investigating Technical and Scale Efficiencies of Australian Universities through Data Envelopment Analysis. Socioeconomic Planning Sciences , 57-80.
3. Berry, J., & Phanthasomchit, S. (2000). Centralized and Decentralized Scheduling Using HL7. Canadian Institute of Health Information.
4. Browlin, W. (1987). Evaluating the efficiency of US air force real-property maintenance activities. Journal of Operational Research society , 127-135.
5. Chiang, W. (2006). A Hotel Performance Evaluation of Taipei International Tourists Hotels- Using Data Envelopment Analysis. Asia Pacicfic Journal of Tourism Research , 29-42.
6. Cooper, W., Deng, H., Gu, B., Li, S., & Thrall, R. (2001). Using DEA to improve the management of congestion in Chinese industries. Socio-Economic Planning Sciences , 227-242.
7. Cooper, W., Lawrence, M., & Tone, K. (2006). Introduction to Data Envelopment Analysis and its usesa. New York: Springer Science Business Media.
8. Ertay, T., & Ruan, D. (2005). Data envelopment analysis based decison model for optimal operator allocation in CSM. European Journal of operational Research , 800-810.

9. Ertay, T., Ruan, D., & Tuzkaya, U. R. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. Information Science , 237-262.

10. Green, L., Savin, S., & Wang, B. (2006). Managing Patient Service in a Diagnostic Medical Facility. Operations Research , 11-25.

11. Guo, M., Wagner, M., & West, C. (2004). Outpatient Clinic Scheduling- A Simulation Approach. Winter Simulation Conference , 1981-1987.

12. Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. IIE Transactions , 800-819.

13. Harper, P. R., & Gamlin, H. M. (2003). Reduced outpatient waiting times with improved appointment scheduling: a simulation modeling approach. OR Spectrum , 207-222.

14. Ho, C.-J., & Lau, H.-S. (1992). Minimizing Total Cost in Scheduling Outpatient Appointment. Management Science , 1750-1764.

15. Hollingsworth, B., & Parkin, D. (1995). The efficiency of Scottish acute hospitals: An application of data envelopment analysis. IMA Journal of Scottish acute hospitals: An application of data envelopment analysis , 161-173.

16. Hooten, M. E. (1990). An effective outpatient appointment system for General Leonard Wood Army Community Hospital. Fort Sam Houston, Tex.: U.S. Army Academy of Health Sciences.

17. Iskan, M. W., Ward, T. J., & McKee, T. C. (1999). Simulating Outpatient Obstetrical Clinics. Winter Simulation Conference , 1557-1563.

18. Jun, J., Jacobson, S., & Swisher, J. (1999). Application of discrete-event simulation in health care clinics: A survey. Operational Research Society , 109-123.

19. Kelton, W., Sadowski, R. P., & Swets, N. B. (2010). Simulation with Arena. New York: McGraw-Hill.

20. Khouja, M. (1995). The Use of Data Envelopment Analysis for Technology Selection. Computers and Industrial Engineering , 123-132.

21. Kirigia, J., Emrouznejad, A., & Sambol, L. (2002). Measurment of Technical Efficiency of Public Hospital in Kenya: using data envelopment analysis. Medical Systems , 39-45.

22. Klassen, K. J., & Rohleder, T. R. (1996). Scheduling Outpatient appointments in dynamic environment. Operations Management , 83-101.

23. Lai, M.-c., Huang, H.-C., & Wang, W.-K. (2010). Designing a knowledge-based system for benchmarking: A DEA approach. Knowledghe-Based Systems , 1-10.

24. Li, X.-B., & Reeves, G. R. (1999). A multiple criteria approach to data envelopment analysis. European Journal of Operational Research , 507-517.

25. Lin, L. C., & Sharp, G. P. (1999). Quantitative and qualitative indices for the plant layout evaluation problem. European Journal of Operational Research , 100-117.

26. Lovell, C., & Pastor, J. (1999). Radial DEA models without inputs or without outputs. European Journal of Operational Research , 46-51.

27. Marinagi, C., Spyropoulos, C., Papatheodorou, C., & Kokkotos, S. (2000). Continual Planning and Scheduling for managing patient test in hospital laboratories. Artificial Intelligence in Medicine , 139-154.

28. Moghadassi, H., Hosseini, A., & Sheikhtaheri, A. (2006). A New Model for the Organizational Structure of the Medical Record Department at Hospitals in Iran, Perspectives in Health Information Management/ AHIMA . Amercian Health Inofrmation Management Association , 1-22.

29. Najmuddin, A., Ibrahim, I., & Ismail, S. (2010). A Simulation Approach: Improving Patient Waiting Time for Multiphase Patient Flow of Obstetrics and Gynecology Department (O&G Department) in Local Specialist Center. WSEAS Transactions on Mathematics , 778-790.

30. Nealon, J., & Moreno, A. (2003). Agent-Based Application in Health Care. In J. Nealon, & A. Moreno, Applications of Software Agent Technology in the Health Care Domain (pp. 3--19).

31. Nickel, S., & Schmidt, U. A. (2009). Process Improvement in Hospitals: A Case Study in a Radiology Department. Quality Management in Health Care , 326-338.

32. Oddi, A., & Cesta, A. (2000). Toward Interactive Scheduling Systems for Managing Medical Resources. Artificial Intelligence in Medicine , 113-138.

33. Oral, M., & Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches. European Journal of Operational Research , 282-294.

34. Ozcan, Y. A. (2008). Health Care Benchmarking and Performance Evaluation. New York: Springer Science + Bussiness Media.

35. Patrick, J., & Putermen, M. (2007). Improving resource utilization for diagnostic service through flexible inpatient scheduling: A method for improving resource utilization. Journal of Operational Research Society , 235-245.

36. Ramanathan, R. (2003). An Introduction to Data Envelopment Analysis. New Delhi: Sage.

37. Shafer, S. M., & Bradford, J. W. (1995). Efficiency Measurment of Alternative Machine Component Grouping Solutions Via Data Envelopment Analysis. IEEE Transactions on Engineering Management , 159-165.

38. Shang, J., & Sueyoshi, T. (1995). A unified framework for the selection of a Flexible Manufacturing System. European Journal of Operational Research , 297-315.

39. Sinuay, Z., & Friedman, L. (1998). DEA and the Discriminant Analysis of Ratios for Ranking Units. European journal of Operational Research , 470-478.

40. Spyropoulos, C. (2000). AI planning and Scheduling in the medical hospital environment. Artificial Intelligence in Medicine , 101-111.

41. Takakuwa, S., & Wijewickrama, A. (2008). Optimizing Staffing Schedule in Light of Patient Satisfaction For the Whole Outpatient Hospital Ward. Winter Simulation Conferences , 1500-1508.

42. Valouxis, C., & Houses, H. (2000). Hybrid Optimization Techniques for workshift and rest assignment of nursing personnel. Artificial Intelligence in Medicine , 155-175.

43. Venkatesh, B. (2006). Technical Efficiency Measurement by Data Envelopment Analysis:An Application in Transportation. Alliance Journal of Business Research .

44. Vermeuleen, L., Bohte, S., Somefun, K., & Poutrae, H. (2007). Multi-agent Pareto Appointment Exchanging in Hospital Patient Scheduling. Service Oriented Computing snd Application , 185-196.

45. Vermeulen, I., Bohte, S., Elkhuizen, S., Bakker, P., & Poutre, H. (2008). Decentralized Online Scheduling of Combination-Appointments in Hospital. International Conference on Automated Planning and Scheduling, (pp. 372-379). Sydney.

46. Vissers, J. (1998). Patient flow-based allocation of inpatient resources: A Case Study. European Journal of Operational Research , 356.

47. Wijewickrama, A. A., & Takakuwa, S. (2006). Simulation Analysis of an Outpatient Deparment of Internal Medicine in a University Hospital. Winter Simulation Conference , 425-432.

48. Wijewickrama, A., & Takakuwa, S. (2008). Outpatient Appointment Scheduling in a Multi Facility System. Winter Simulation Conferences , 1563-1571.

49. Wijewickrama, A., & Takakuwa, S. (2005). Simulation Analysis of Appointment schduling in an Outpatient Department of Internal medicine. Winter Simulation Conference , 2264-2272.

50. Yang, T., & Kuo, C. (2003). A hierarchical AHP/DEA methodology for the facility layout design problem. European Journal of Operational Research , 128-136.

51. Zhang, J., Dharmadhikari, N., & Song, D. (2010). Literature Analysis on Centralized and Decentralized Scheduling. Fargo: Department of Industrial and Manufacturing Engineering.

52. Zhang, J., Dharmadhikari, N., & Song, D. (2009). Literature Review on Centralized and Decentralized Scheduling. Fargo: Department of Industrial and Manufacturing Engineering.

53. Zhang, J., Dharmadhikari, N., & Song, D. (2010). Summary of qualitative and quantitative questionnaire findings in centralized versus decentralized scheduling system. Fargo: Department of Industrial Engineering and Manufacturing, NDSU.

54. Zhang, J., Gonela, V., & Aslani, N. (2010). Develop a decision tool for a hospital/clinic setting to decisde whether to schedule in a Centralized or Decentralized manner. Fargo: Department of Industrial and Manufacturing Engineering, NDSU.

55. Zhang, J., Gonela, V., & Aslani, N. (2011). Development of Centralized, Decentralized and Hybrid Scheduling Model. Fargo: Department of Industrial and Manufacturing Engineering, NDSU.

56. Zhang, Y., & Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia Sweden and New Zealand. Journal of Productivity Analysis , 187-204.