International Journal of Operations Research

International Journal of Operations Research Vol. 15, No. 1, 15−28 (2018)

A Simulation Study of Workforce Management for a Two-Stage Multi-Skill Customer Service Center

Chia-Hung Wang[1,2]* and Peiqiao Su[1]

[1]College of Information Science and Engineering, Fujian University of Technology,
Fuzhou City, Fujian Province 350118, China

[2]Fujian Provincial Key Laboratory of Big Data Mining and Applications,
Fuzhou City, Fujian Province 350118, China

Received February 2018; Revised March 2018; Accepted March 2018

Abstract: This paper considers the parameter optimization of a two-stage multi-skill customer service center, which provides e-commerce services to customers and bears the major operating cost in hiring service agents. Based on the customer flow in the two-stage mixed queueing system, a simulation model is developed to determine the system performance of interest. We present a sensitivity analysis to achieve better savings in the staffing cost and waiting time at a guaranteed service level. A series of simulation experiments are conducted via an Arena simulation platform to figure out the optimal system configurations. The contribution of our works is to provide a decision-making tool for workforce managers to evaluate the performance of the studied customer service centers.

Keyword — Workforce management, Arena simulation, queueing system, customer service center, sensitivity analysis.

1. INTRODUCTION

China's customer service industry has been rapidly expanding with the vast development of e-commerce markets in recent years. According to the survey of Zero Power Intelligence Group (2015), the staffing number reached 961.1 thousand in the whole mainland China's customer service centers, and China's call center market was reached 110.5 billion RMB. Modern customer service centers are challenged with multitude types of calls, including telephone, Internet, WeChat/QQ, mobile devices, fax, e-mail, etc. The customer service centers have become important connections between e-commerce enterprises and customers.

Workforce management is critical in the customer service centers. Taking a large e-commerce company JD Mall in China as an example, thousands of customers' (online shopping) requests are handled by hundreds of service agents every day. With the growing number of agents working in the service sector, labor costs constitute a substantial part of the business expenses. According to the survey of Zero Power Intelligence Group (2015), the cost of hiring employee is about 60% to 70% of the operating costs in a customer service center. Besides, the number of service agents is required to be planned in order to keep the customer satisfaction level over a predefined level. An inadequately sized workforce can lead to long waiting times and low service levels. This can be avoided by scheduling a sufficiently large number of service agents. However, it is undesirable for managers to assign too many agents due to high employee salaries. So, it is important to find a good match between the predicted workload and the allocation of labor resources.

In this paper, we deal with the workforce management problem for a two-stage multi-skill customer service center at a low cost while attaining a satisfactory service level for incoming requests. Managing an appropriate staffing level becomes a leading issue because the labor costs can reach the most of total operating costs in the customer service center, where the service level is guaranteed at least 80%. Low staffing levels could lead to long waiting times and low service level, but it is costly to maintain high staffing levels in a customer service center. Thus, we need a proper agent allocation plan with low staffing cost while meeting the targeted service level and keeping agent utilizations at reasonable levels. In our study, several workforce management strategies will be explored for staffing in a multi-skill customer service center, where service agents belonging to different skill sets have different knowledge levels in handling particular types of incoming customers' requests.

We develop an Arena simulation model for understanding and assessing the likely impact of system configurations on the system performance. Our Arena simulation model is a collection of modules, including data modules, logic modules and process modules. Each module contains all of the model parameters, logic, and

*Corresponding author's e-mail: jhwang728@hotmail.com

http://doi.org/10.6886/IJOR.201803_15(1).0002

animation necessary to describe its specific portion of the studied system. In the customer service center of our study, the decision variable is the number of service agents to be allocated to each type of skill sets. With simulation methodology, we will conduct the what-if analysis, and a series of scenarios will be generated to evaluate the performance of the proposed system configuration that directly affects the economical profitability and the operational efficiency.

Achieving a desired balance between the service level and operational efficiency is a key challenge in managing the studied customer service centers, where the service level is targeted as 80% of customers can receive service in 20 seconds. A natural consequence of customers' dissatisfaction with the system performance is that customers may lose their patience and abandon the queue. The (near) optimal agent allocation schemes will be determined through analyzing the numerical results in our simulation experiments.

The main contribution of this study is to perform a sensitivity analysis for managing a two-stage multi-skill customer service center. With our Arena simulation model, it provides an efficient quantitative approach to evaluate the system performance of interest. The objective of the proposed sensitivity analysis is to explore the (near) optimal agent allocation schemes which achieve the minimal staffing costs while maintaining a guaranteed 80% service level and low waiting time. In addition, several managerial insights will be figured out for optimizing the parameter settings of the studied customer service center.

The remainder of this paper is organized as follows: In Section 2, we introduce a literature review on the usage of simulation methodology for solving workforce management problems. In Section 3, a simulation model will be developed for studying the two-stage multi-skill customer service center via the Arena simulation software. In Section 4, we conduct a series of sensitivity analysis to demonstrate the effects of varying the staffing numbers in different service types on the system performance. Finally, the concluding remarks and our suggestions for future works are summarized in Section 5.

2. LITERATURE REVIEW

Computer simulation methods have been used to analyze the customer flow and resource allocation in the service industries over recent decades. The virtual queue lines occur in a customer service center when there is no service agent available to handle a customer's request. Through the simulation experiments, we can translate the gross data in several important performance measures of a service center, e.g., the service levels, the number of agents and the corresponding costs, etc. As mentioned in Bouzada (2009), the applications of simulation have become an essential element in determining the optimal way to serve customers efficiently and effectively.

There is a growing focus in literature on simulation studies to help service providers/organizations in expanding the operational efficiency, enhancing the level of service, and reducing costs. For example, Mehrotra and Fama (2003) provided an overview of call center simulation models, where they highlighted typical inputs and data sources, modeling challenges, and key model outputs. Wallace and Saltzman (2005) provided a comparison of the pros and cons of using two simulation programming methods (the C language and the Arena software package) when describing the modeling of a skill-based routing call center. An inbound call center of a city-gas company was simulated in Takakuwa and Okada (2005) to examine the proper target of the service level and to explore the optimal number of agents. Bouzada (2009) suggested a simulation approach for the dimensioning of a service handling capacity, and discussed its adequateness to complex operations in a large Brazilian call centers company which were in detriment of analytical methods (such as Queue Theory). Wang and Zhu (2017) developed an Arena simulation model to study the impact of customers' impatient behaviors (in terms of balking and reneging) on the performance of a queueing system in the contact centers. Manoel et al. (2017) provided a systematic literature review on the use of simulation to develop IT helpdesk services. Greasley and Smith (2017) suggested the use of discrete-event simulation to model workforce staffing scenarios at a police communications center, which undertakes a vital role in receiving and processing emergency and non-emergency telephone calls from the public and other agencies. Leva et al. (2017) adopted a simulation framework to study different system operating scenarios for a public contact center in the northern Italy, which furnished health/social information to citizens and provided different typologies of answers to users' requests. As mentioned in Munoz and Brutus (2013), simulation represents a good alternative not only to determine an adequate number of service agents required to achieve the operational targets, but also as a way to visualize the relationship among the competing objective scenarios.

Determining an adequate workforce size to achieve the target of service level is a complex and important decision for managers. For dealing with the workforce management problems, Atlason et al. (2008) studied the staffing problem in an inbound call center, while maintaining an acceptable service level in multiple time periods. Bhulai et al. (2008) determined the staffing levels and used the outcomes as input for shift scheduling problem in multi-skill call centers. Sencer and Ozel (2013) showed the advantages of simulation modeling on generating a more fast, reliable and efficient decision-making environment for call centers in Turkey. Wang and Hung (2016) studied the optimal design for staffing problems in the call centers with a single function or double functions. Erekat et al. (2017)

developed a simulation model with the Arena software to test various staffing alternatives for an outpatient access center. In the case study of Greasley and Smith (2017), they addressed the conflict between the need to reduce cost and the requirement to meet national standards in terms of a timely response to customer calls, and provided an assessment of strategies that aim to reduce staffing cost whilst maintaining service levels in a police operation.

The service level of a given contact center is typically quantified in terms of some congestion or performance measures. For example, Koole and Mandelbaum (2002) suggested a managerial focus on the customers' abandonment rate and waiting time. For a large Chilean call center, Munoz and Brutus (2013) illustrated the trade-off among the number of multiple-skill operators and the bi-criteria target (service level and abandonment) through a discrete-event simulation model. In a service-based queueing system involving human servers, Wang and Zhou (2017) investigated the impact of queue configuration on the service time of human servers from the empirical analysis, and demonstrated that pooling could have an indirect negative effect on service time through its impact on queue length. Using behavioral experiments, Shunko et al. (2018) studied the impact of queue design on worker productivity in service systems that involve human servers, and they found that the single-queue structure slowed down the servers.

The discussions above show that there are numerous studies that consider simulation as an efficient methodology in the agent allocation plans. This paper differs from most of the previous studies by allowing labor resource allocation in a two-stage multi-skill environment. The proposed simulation model in our study will be used in measuring the system performance and in testing whether some changes would be able to improve the studied system before its implementation.

3. PROBLEM DEFINITION AND MODELING

In this section, a simulation model is developed for the two-stage multi-skill customer service center using Arena software and then used as a baseline model to evaluate all proposed scenarios. The motivation for the choice of Arena simulation software in our study is that Arena offers a user-friendly modeling environment that can be integrated with a database system, as well as graphical user interfaces for data input and reporting. In addition, there are standard modules for common functions and properties, and it also integrates Visual Basic for Applications into its product architecture so that we can create our own utility tools and custom interfaces. For more detailed information, interested readers may refer to Rockwell Automation (2006) and Altiok and Melamed (2010).

3.1 Workforce Management for a Two-stage Multi-skill Customer Service Center

Figure 1 illustrates the process flow chart of a two-stage multi-skill customer service center and its nature of work. The basic process of the presented simulation model is to generate a stream of arriving customers' requests, assign them to queueing lines, route them to a specific type of service agents, and terminate them. When the customers' requests arrive at the customer service center, the customer gets a busy signal and has to wait in the queue of the related agent group if all of the service agents are busy; otherwise, his/her request is handled by one of the agents who provide the corresponding services. There are three service types (skill sets) of agents within the customer service center. A service type of agents represents a group of agents who have the same skill sets, that is, an agent group is a set of identical agents from a simulation modeling perspective. Agent service types and their skill/knowledge levels in fulfilling different type of requests can be defined in the simulation model.

As shown in Figure 1, when the customers' requests arrive at the system, they will be classified as service type I or service type II according to the type of customer's needs, and then enter the corresponding queueing lines. When there is at least one available service agent, the incoming request can be handled immediately without waiting for service. Otherwise, when there is no available service agent, the requests in the same service type wait in the queueing lines according to the order of their arrivals until (at least) one service agent is available. The queue lengths are set as unlimited, and we assume the number of potential customers is also unlimited.

When the requests wait in the queue, the system will determine whether the queue length exceeds a preset value (acceptable queue length) and then prompt message to waiting customers. Providing the delay time information to the customers is one of the managerial approaches to achieve the service level and the operational efficiency of customer service centers. For further discussions, interested readers may refer to the empirical study in Akşin et al. (2017), where the impacts of delay announcements on callers' abandonment behavior and the performance of a call center were analyzed. If the waiting queue is too long, the customers' requests in the queue may be canceled (and then leave the system directly) with an abandonment probability. For the requests waiting in the queue of service type I, it may be abandoned with a probability P1. For the waiting requests of service type II, it may be canceled without service with a probability P2.

After the request is processed in the service type I or type II, there is a transfer probability P3 that the customers' requests may further need an advanced service (service type III). That is, the requests handled by service agents in type I or type II may be completed (and leave the system) with a probability 1- P3, and may be transferred to the queueing lines for advanced service with a probability P3. After completing one service request, the customer service center will record the satisfaction level of that customer, which will be used as an indicator of system performance.

In our study, the Service Level equals the number of satisfied requests divided by total number of served requests, where the satisfied request is defined as the requests that can be handled by a service agent in 20 seconds. The Abandonment Rates are defined as the percentage of customers' requests canceled before reaching a service agent. The Staffing Utilization Rate is calculated as the proportion of the average service time of service agents compared to their entire working time in predefined time intervals, which can be used to discriminate the rationality of agent allocation schemes. The Average Waiting Time is defined as the average time that all requests wait in the queueing lines before getting a service during the simulation system operates.

The decision variables in this study are the numbers of service agents to be allocated in three service types. The service level of the studied customer service center is targeted as that the number of customers' requests which can get service within 20 seconds is required to be over 80% of the total requests. The objective of our study is to find an optimal agent allocation at a low staffing cost while meeting the targeted service level for customers' requests.

3.2 An Arena Simulation Model

Based on the process flow chart illustrated in Figure 1, we develop the simulation model with Arena software. As shown in Figure 2, the flow of customers' requests can be monitored via the visualization of the Arena simulation model to make sure the proposed system works as expected.

The process of this simulation model is described as follows. The Create module is a starting point of our simulation model. We use the Create module to randomly generate the customers' requests, and the number of customers in the Create module is set as unlimited. The customers' requests are dynamic objects named as Entities in the Arena simulation system, where the Entity module is used to define an entity and assign its attributes. The arriving requests enter the transmission channels in the customer service center, and then select service types they need. In the present model, there are three types of service agents.

There is a preset probability that incoming requests will be assigned as the service type I, and the others will be assigned as the service type II. We determine the service type (type I or type II) of the incoming requests by using a Decide module named Select Service in Figure 2, where the proportion/probability can be set to operate the selection of two service types. An attribute (service type) is added to a customer's request to make it characterized, and different type of requests use different attributes to distinguish each other. Depending on the attributes of customers' requests, we separate them to different queue lines by using the Decide module, and then assign them to a corresponding queue.

The simulation system determines whether the current queue length is greater than a given value (acceptable queue length) through a Decide module named as Judge Number module. If the current queue length exceeds a preset value, the waiting entities may abandon the queue with a given probability. Otherwise, they go through a series of specific operations in the corresponding service type. In our Arena simulation model, the Resource module is used to set the service agents in the corresponding service type, and a Resource represents a service agent. If all of the Resources in a specific service type are seized, the entity waiting in the corresponding queue may abandon or stay in the queue to wait for service.

After the entities complete the service in type I (or type II), they will enter a Decide module named Transition Probability 1 (or Transition Probability 2), and determine whether the corresponding entity needs to be transferred to the queueing lines in service type III. A transfer probability is set in these two modules according the statistics recording the proportion of customers' requests transferred for an advanced service. Through those Decide modules named as Judgment Satisfaction modules, we can determine whether the waiting time of an entity exceeds the predetermined target (20 seconds). The Record modules for satisfaction records are used to calculate the service levels. The Variables in the Arena platform can be used to calculate the statistics of system performance, such as average waiting time, service level, resource utilization, etc.

The Assign module is used to assign a numerical value to the Variables defined by users in the Arena simulation system, and the Record module equipped with several statistical functions can be used to record the passing entities. The Queue module is used to define the queuing rules in the simulation system, where the requests in the same service type are served based on a First-Come-First-Served rule. The Dispose modules are used to delete those entities created by the Arena simulation system.

In the Arena simulation model developed here, we take the arrival rates of customers' requests, average service (handle) times, and agent allocation plan as inputs. The average arrival rates and service times can be obtained from historical data for each service type, such as the statistical analysis of the input data provided in Sencer and Ozel

(2013).

The decision variables in our model are the numbers of service agents to be allocated to three types of skill sets, that is, the numbers of agents for service type I, service type II and advanced service (type III). We set the available number of agents in each service type with the Process modules, and those possible skill-set-based agent allocation schemes will be explored for the (near) optimal number of agents assigned to each service type.

Our Arena simulation model can be used as a flexible tool to evaluate the performance of a customer service center by defining various comparison scenarios, which will be tested and discussed in the next section. In the simulation experiments, we will report the system performance of those agent allocation schemes, including the service level, average waiting time, and staffing utilization rate.

4. NUMERICAL RESULTS

In this section, we conduct the sensitivity analysis to study the managerial effect of the interested model parameters on the system performance of the proposed Arena simulation model. Through a series of simulation experiments, we will observe the effect on the system performance by varying the staffing numbers in three service types, respectively. Several feasible solutions to the staffing problem will also be explored in the numerical results. The objective of the presented sensitivity analysis is to determine the (near) optimal agent allocation schemes with low staffing costs while maintaining a guaranteed 80% service level and low waiting time.

4.1 Experimental Parameter Settings

In our simulation experiments, the Arena model parameters are set according to the empirical data from a call center in a logistics company taken in Su and Zhao (2015). The average arrival rate to the studied customer service center is 84.44 customers per minute, where the arrival rate to service type I is 27.98 customers per minute, and the arrival rate to service type II is 56.46 customers per minute. When incoming customers choose service type, there is 33.14% proportion of customers to service type I, and the others (66.86% of incoming customers) will go to service type II. The abandonment probabilities P1 and P2 are 0.05%. The transfer probability P3 to advanced service (type III) is 7.31%, that is, the arrival rate to advanced service type III is 6.17 customers per minute. The inter-arrival time is exponentially distributed, and the customer source is assumed to be infinity.

The average staffing cost for hiring one service type I agent is 12 RMB per hour. The average staffing cost for service type II agents is 15 RMB per person per hour, and the average staffing cost for advanced service (type III) agents is 20 RMB per person per hour. The average service time for type I is 1.74 minutes, the average service time for type II is 0.81 minutes, and the service time for advanced service type III is 0.89 minutes. The service times of customers follow exponential distributions, and the waiting space for all service types is unlimited. The service level is determined by the proportion of satisfied customers, and our target for the service level of the studied customer service center is above 80%.

In our simulations, the warm-up time was set for 4 hours and the running time was set for 7 days, in order to ensure a sufficiently long running time and a sufficiently large number of runs to reduce the experimental error. Each simulation experiment was repeated 10 times independently, and the numerical results provided in this paper are obtained by averaging those values of 10 experimental data. Our numerical experiments are run through Arena simulation software version 14.00 on the PC platform with Intel Core i5-2520M (2.5 GHz) and 8 GB RAM.

4.2 Sensitivity Analysis

Here, we study the effects of varying the staffing numbers on the system performance of our interest. The staffing number in service type I varies from 45 to 60. The staffing number in service type II varies from 42 to 53. The staffing number in advanced service type III varies from 4 to 12. These combinations result in 16*12*9=1,728 simulation scenarios.

In order to reduce the feasible solution region, we conduct a sensitivity analysis of changing the staffing number on the service level. Firstly, we investigate the impact of varying the staffing numbers in service type I on the service levels, and the numerical results are depicted in Figure 3. It can be observed that the curves are increasing and S-shaped. It illustrates that the satisfaction of customers in service type I and the service level of overall system are positively correlated to the number of agents in service type I. Besides, the degree of growth decreases after the data point where the staffing number equals 48, and the growth rate goes down to nearly zero after the staffing number reaches 54. Therefore, for efficiently determining the (near) optimal schemes, the searching space for the number of agents in service type I can be set as S1={49, 50, 51, 52, 53, 54}.

Secondly, we observe the effect of varying the staffing numbers in service type II on the service levels. From the numerical results depicted in Figure 4, it shows that both the service levels of service type II and overall system are increasing when we increase the number of agents in service type II. In order to achieve the targeted service level 80%, we observe that the staffing number should be given at least 45 in service type II. Meanwhile, the cost investment for hiring agents of service type II becomes worthless after the data point where the staffing number equals 49. Hence, for determining the (near) optimal number of agents in service type II, the searching set can be reduced as S2={45, 46, 47, 48, 49}.

Similarly, we study the sensitivity analysis of the system performance when increasing the number of agents for advanced service (type III). In Figure 5, when varying the staffing number in advanced service from 4 to 12, it depicts those numerical results obtained in the Arena simulation experiments. It can be found that the targeted service level would not be achieved if the staffing number is less than 7, and the degree of growth in the customers' satisfaction is not effective when the staffing number achieves 9. Therefore, we set the searching points for the staffing number in service type III as S3={7, 8, 9}.

Through the above sensitivity analysis, we can reduce the original solution space (1,728 scenarios) into smaller solution space (6*5*3=90 scenarios). Moreover, taking the targeted service level 80% as the boundary line, we can split each searching set S1, S2, and S3, individually. For example, the searching set S1 can be divided as {49, 50} and {51, 52, 53, 54} when splitting it with 80% service level for customers' requests of type I. Similarly, the searching set S2 can be divided as {45, 46} and {47, 48, 49} when splitting it with 80% service level for those requests of type II. We select those integers whose performance less than 80% service level in set S1 (S2) and those numbers performing more than 80% service level in set S2 (S1), respectively. After combining them with the numbers in the set S3, it can be reduced to be only 45 possible schemes such that the overall system achieves nearly 80% service level. Table 1 lists all of 45 possible schemes (simulation scenarios) and the corresponding staffing costs. For each possible scheme in Table 1, we conduct the simulation experiments with the setting of staffing numbers for service type I, type II and type III in the Arena model. Table 2 summarizes the service levels determined from the simulation experiments for 45 possible schemes. The service levels and total staffing costs of 45 agent allocation schemes are depicted in Figure 6.

4.3 A Strategy for Selecting Better Agent Allocation Schemes

From 45 possible schemes in Figure 6, we select the schemes with both service level more than 80% and the staffing cost less than 1,458 RMB/hour. There are 4 schemes that are satisfied with the requirements, i.e., Scheme 4, Scheme 10, Scheme 28, and Scheme 43. The system performances of these four feasible schemes are summarized in Table 3.

When considering the minimum staffing cost, Scheme 10 may be the best choice of four schemes. However, when considering the service level of overall system, we get another preference order:

Scheme 43 > Scheme 10 > Scheme 4 > Scheme 28.

In the following, we explore a better preference order to balance a trade-off between these two targets by conducting a comprehensive analysis of the performance evaluations for four feasible schemes in Table 3.

For Scheme 4, although it makes the targeted service level of overall system, the satisfaction level of service type I is very low (only 59.57%) while the satisfaction level of type II achieves 93.7%. The average waiting time in service type I is very large, which easily leads to dissatisfaction with the service of the enterprise and a large loss of customers. Although the total staffing cost of Scheme 4 is low, the gap of performance in two service types is too large. So, this scheme is obviously not desirable.

The staffing cost of Scheme 10 is the lowest of four schemes. Compared to Scheme 4, the satisfaction levels of three service types in Scheme 10 is more balanced, and the service level of overall system is higher than Scheme 4.

For Scheme 28, the overall system performance is slightly lower than Scheme 10, and the cost is also higher than Scheme 10. The satisfaction levels of all service types in Scheme 28 are more balanced relative to Scheme 4. For considering the staffing utilization, the utilization rate of service type I is too low, which easily leads to a waste of workforce resource in service type I.

For Scheme 43, the satisfaction levels of type I and II is the most balanced among 4 schemes, and the service level of overall system is also the highest. Besides, the average waiting time and staffing utilization are also well under control in Scheme 43. In terms of the overall performance, Scheme 43 is the best of the four schemes, but it is also the most costly.

From the above discussions, we suggest a cost-effective preference order as follows:

Scheme 10 > Scheme 43 > Scheme 28 > Scheme 4.

Scheme 10 is the most cost-effective because it can save more staffing costs and hence creates more profits under the premise of not losing customers. For the e-commence enterprises that are already in a steady state of business development, we can choose Scheme 10 as the optimal solution to allocate service agents in their customer service centers. Nevertheless, we suggest Scheme 43 as the optimal solution for those e-commerce enterprises which intend to accumulate customers and lay a good reputation. Through investing more in hiring employee, the company could build a better customer service center to retain customers and accumulate a good foundation for the future development.

5. CONCLUSIONS

In this research, we determined (near) optimal agent allocations for a two-stage multi-skill customer service center. A strategy for selecting cost-effective schemes is presented to balance the trade-off between low staffing cost and high service level. We developed an Arena simulation model to evaluate the effect of what-if scenarios before implementing them. The sensitivity analysis has been conducted through a series of numerical experiments with different model parameters, such as varying the staffing numbers in service type I, type II and advanced service (type III). From the experimental data in our study, we find that there are several significant trends in the change of the service level when varying the number of service agents. The presented sensitivity analysis is helpful in exploring better workforce management for the customer service center with a targeted service level.

In the practical applications, the proposed simulation model can be further adjusted by the workforce managers in order to analyze and test the impacts of the configuration changes of their interest. Our Arena simulation model is a simplified representation of reality, and can be used to figure out the key relationships and dynamics in the studied customer service centers. In addition, it can also be used to evaluate the staffing levels and the likely impact of changes on the interested performance indicators before implementing them on the actual/real-world system. In the future works, the obtained agent allocation for three service types could be used as the input of an efficient algorithm for determining the shift scheduling in daily operations.

ACKNOWLEDGMENT

This research was supported in part by Fujian Provincial Department of Science and Technology, China, under Grant No. 2016J01330, and the Education Department of Fujian Province, China, under Grant No. JAT160337.

REFERENCES

1. Akşin, Z., Ata, B., Emadi, S.M., and Su, C.-L. (2017). Impact of Delay Announcements in Call Centers: An Empirical Approach. Operations Research, 65(1): 242-265.
2. Altiok, T. and Melamed, B. (2010). Simulation Modeling and Analysis with ARENA, Academic Press.
3. Atlason, J., Epelman, M.A., and Henderson, S.G. (2008). Optimizing Call Center Staffing using Simulation and Analytic Center Cutting-Plane Methods. Management Science, 54(2): 295-309.
4. Bhulai, S., Koole, G., and Pot, A. (2008). Simple Methods for Shift Scheduling in Multiskill Call Centers. Manufacturing & Service Operations Management, 10(3): 411-420.
5. Bouzada, M.A.C. (2009). Dimensioning a Call Center: Simulation or Queue Theory? Journal of Operations and Supply Chain Management, 2(2): 34-46.
6. Erekat, A., Al-Shaebi, A., Alhaider, A., Khasawneh, M., and Khasawneh, A. (2017). Optimizing Outpatient Access Center Staffing Using Discrete Event Simulation: A Case Study. Proceedings of the 2017 Industrial and Systems Engineering Conference, 1918-1924.
7. Greasley, A. and Smith, C.M. (2017). Using Activity-Based Costing and Simulation to Reduce Cost at a Police Communications Centre. Policing: An International Journal of Police Strategies & Management, 40(2): 426-441.
8. Koole, G. and Mandelbaum, A. (2002). Queueing Models of Call Centers: An Introduction. Annals of Operations Research, 113: 41-59.
9. Leva, A.D., Sulis, E., and Vinai, M. (2017). Business Process Analysis and Simulation: The Contact Center of a Public Health and Social Information Office. Intelligent Information Management, 9: 189-205.
10. Manoel, L.G., Bouzada, M.A.C., and Alencar, A.J. (2017). Computer Simulation Improving the IT Helpdesk Problem Management: A Systematic Literature Review. International Business Management, 11(1): 68-77.
11. Mehrotra, V. and Fama, J. (2003). Call Center Simulation Modeling: Methods, Challenges and Opportunities. Proceedings of the 2003 Winter Simulation Conference, 135-143.
12. Munoz, D. and Brutus, M.C. (2013). Understanding the Trade-offs in a Call Center. Proceedings of the 2013 Winter Simulation Conference, 3992-3993.
13. Rockwell Automation (2006). Arena Contact Center User's Guide, Rockwell Automation Technologies.
14. Sencer, A. and Ozel, B.B. (2013). A Simulation-based Decision Support System for Workforce Management in Call Centers. Simulation: Transactions of the Society for Modeling and Simulation International, 89(4): 481-497.

15. Shunko, M., Niederhoff, J., and Rosokha, Y. (2018). Humans Are Not Machines: The Behavioral Impact of Queueing Design on Service Time. Management Science, 64(1): 453-473.
16. Su, Q. and Zhao, F. (2015). Study on the Staffing Algorithm for Multi-skill Call Center. Systems Engineering – Theory and Practice, 35(1): 239-246.
17. Takakuwa, S. and Okada, T. (2005). Simulation Analysis of Inbound Call Center of a City-Gas Company. Proceedings of the 2005 Winter Simulation Conference, 2026-2033.
18. Wallace, R.B. and Saltzman, R.M. (2005). Comparing Skill-Based Routing Call Center Simulations using C Programming and Arena Models. Proceedings of the 2005 Winter Simulation Conference, 2636-2644.
19. Wang, C.-H. and Hung, M.-H. (2016). Simulation Model and Optimal Design for Call Center Staffing Problems. Advances in Intelligent Systems and Computing, 536: 103-110.
20. Wang, C.-H. and Zhu, B. (2017). Sensitivity Analysis of a Simulation Model for Managing Contact Centers with Impatient Customers. Advances in Intelligent Systems Research, 153: 307-311.
21. Wang, J. and Zhou, Y.-P. (2017). Impact of Queue Configuration on Service Time: Evidence from a Supermarket. Management Science, Published online in Articles in Advance 19 Jul 2017. Available at https://doi.org/10.1287/mnsc.2017.2781
22. Zero Power Intelligence Group (2015). Research and Consultation Report of Panorama Survey and Investment Strategy on China's Call Center Industry in Year 2015-2020, Beijing, China, 2015. Available at http://www.chinairn.com

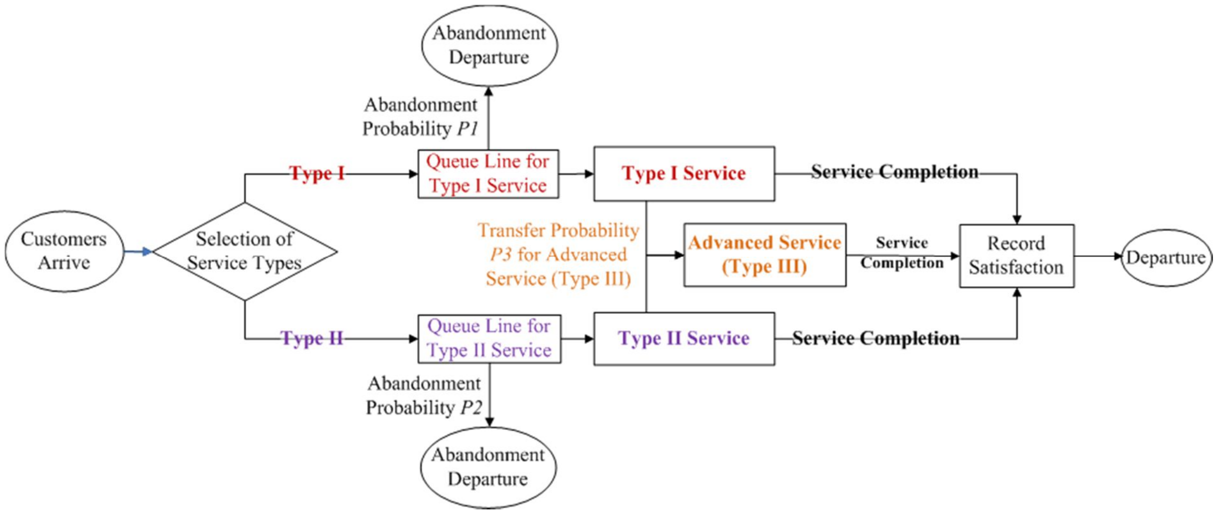

Figure 1. A flow chart for a two-stage multi-skill customer service center.

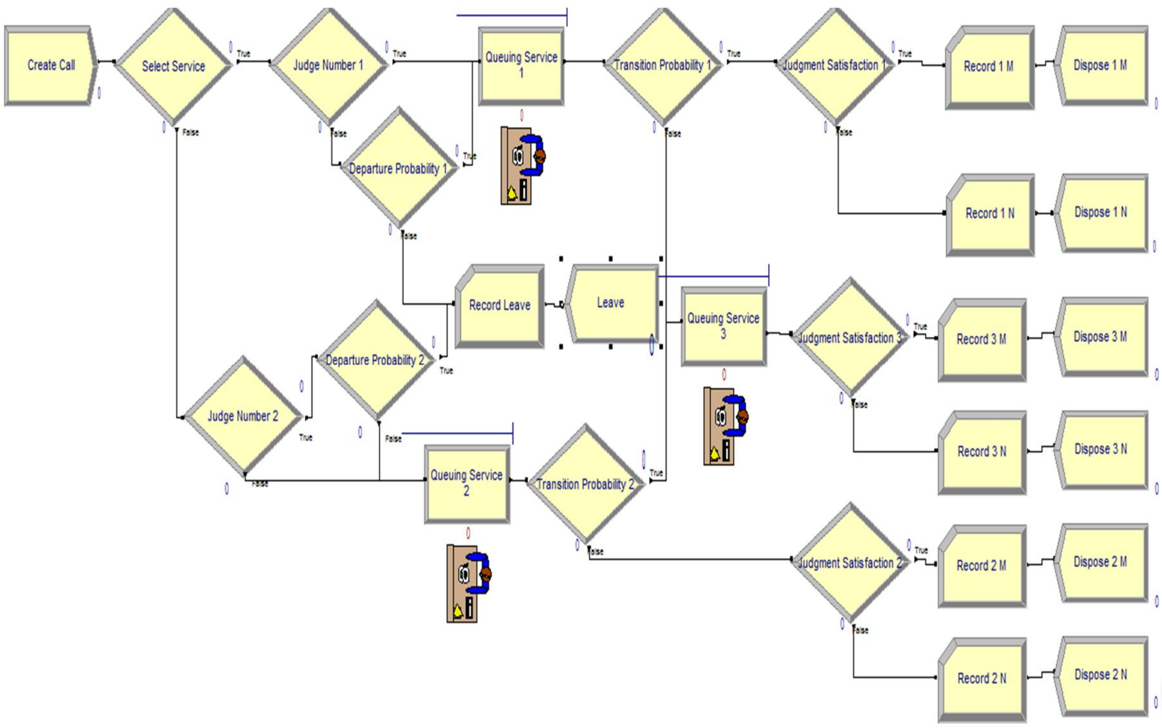

Figure 2. Arena simulation model for a two-stage multi-skill customer service center.

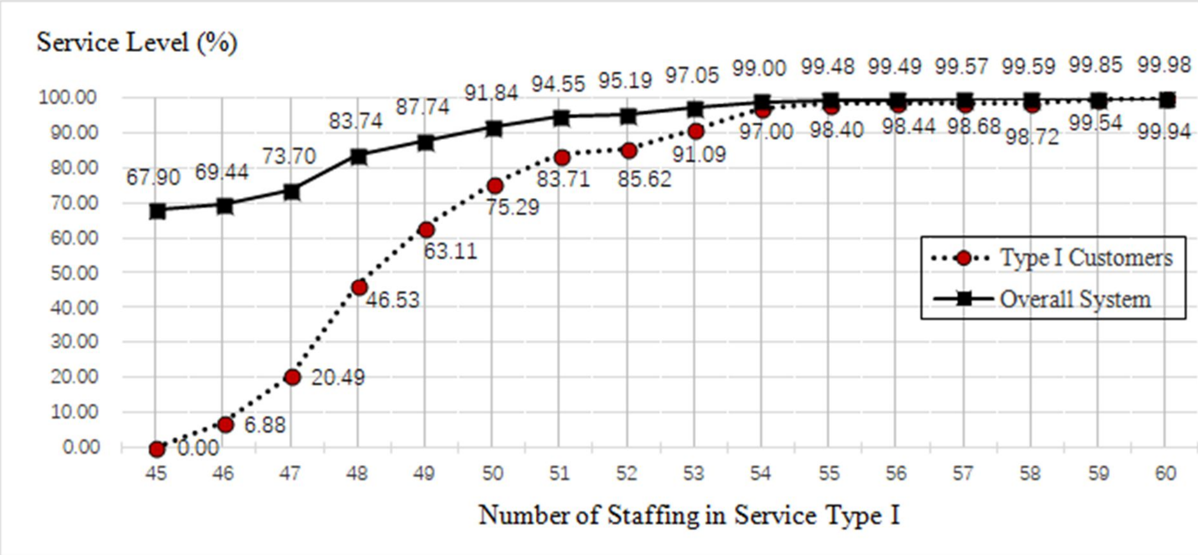

Figure 3. The impact of varying the number of agents in service type I on the service level.

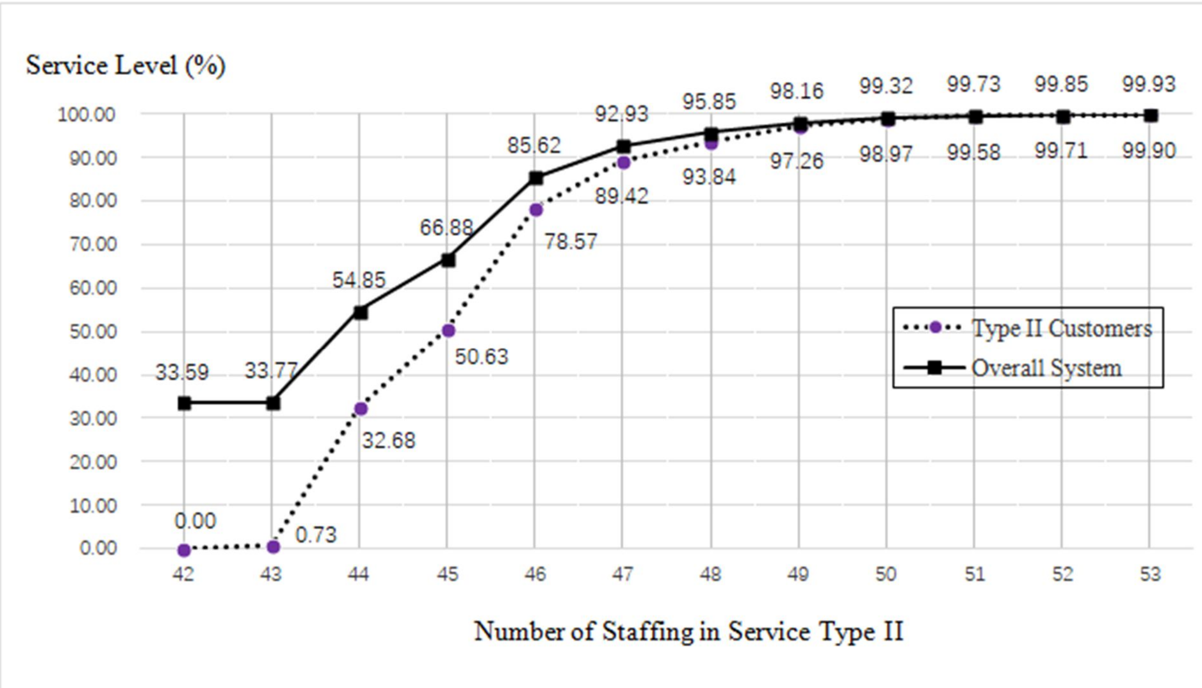

Figure 4. The impact of varying the number of agents in service type II on the service level.

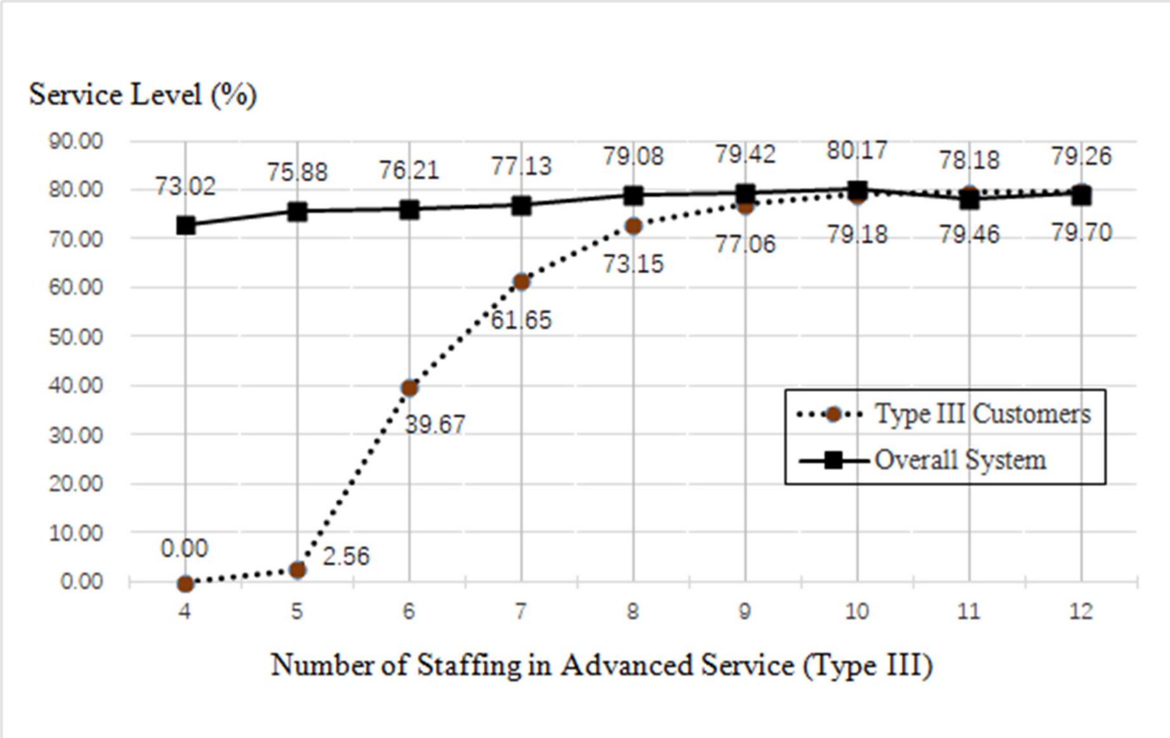

Figure 5. The impact of varying the number of agents in advanced service (type III) on the service level.

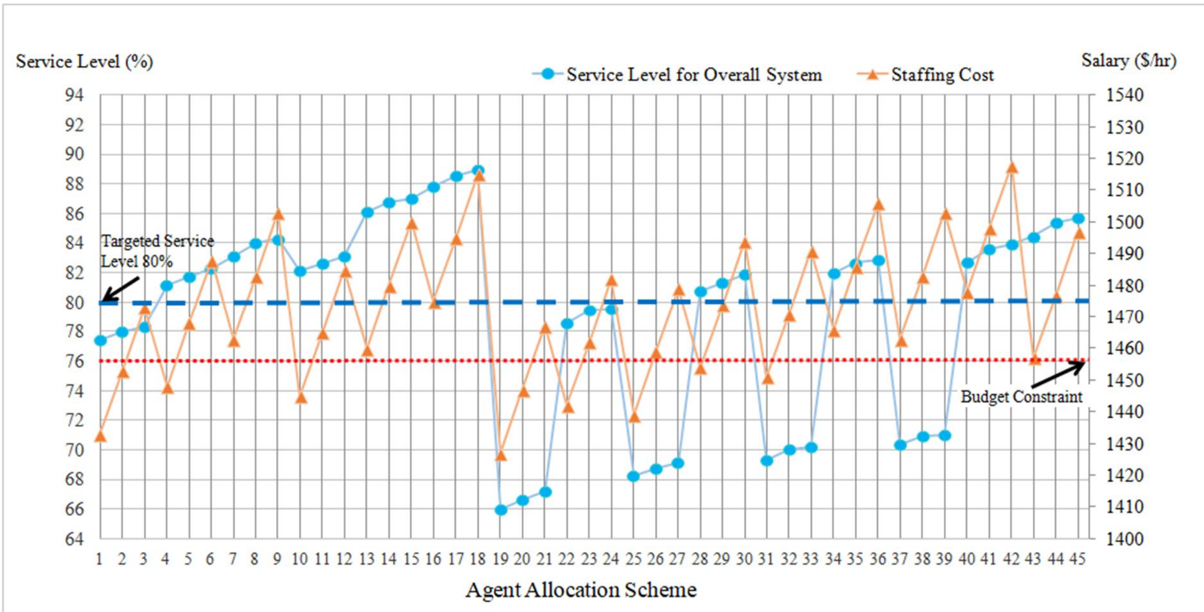

Figure 6. The service level and staffing cost for 45 agent allocation schemes.

Table 1. The Numbers of Agents in Three Service Types and Staffing Costs for 45 Possible Schemes

Scheme	Staffing Number in Service Type I	Staffing Number in Service Type II	Staffing Number in Service Type III	Total Cost
1	49	47	7	1433
2	49	47	8	1453
3	49	47	9	1473
4	49	48	7	1448
5	49	48	8	1468
6	49	48	9	1488
7	49	49	7	1463
8	49	49	8	1483
9	49	49	9	1503
10	50	47	7	1445
11	50	47	8	1465
12	50	47	9	1485
13	50	48	7	1460
14	50	48	8	1480
15	50	48	9	1500
16	50	49	7	1475
17	50	49	8	1495
18	50	49	9	1515
19	51	45	7	1427
20	51	45	8	1447
21	51	45	9	1467
22	51	46	7	1442
23	51	46	8	1462
24	51	46	9	1482
25	52	45	7	1439
26	52	45	8	1459
27	52	45	9	1479
28	52	46	7	1454
29	52	46	8	1474
30	52	46	9	1494
31	53	45	7	1451
32	53	45	8	1471
33	53	45	9	1491
34	53	46	7	1466
35	53	46	8	1486
36	53	46	9	1506
37	54	45	7	1463
38	54	45	8	1483
39	54	45	9	1503
40	54	46	7	1478
41	54	46	8	1498
42	54	46	9	1518
43	51	47	7	1457
44	51	47	8	1477
45	51	47	9	1497

Table 2. Service Levels for 45 Agent Allocation Schemes

Scheme	Service Level (%)			
	Service Type I	Service Type II	Service Type III	Overall System
1	59.96	87.96	62.34	77.41
2	59.96	87.85	70.70	77.98
3	59.82	87.71	75.92	78.28
4	59.57	93.70	66.13	81.16
5	59.58	93.59	74.84	81.73
6	59.99	93.65	79.65	82.27
7	59.82	96.75	66.33	83.12
8	60.14	96.69	77.33	84.01
9	60.10	96.86	80.07	84.29
10	74.45	87.88	66.34	82.16
11	74.16	87.78	74.09	82.60
12	74.25	87.78	80.86	83.12
13	74.29	93.90	69.30	86.09
14	74.20	93.61	81.05	86.73
15	74.00	93.75	84.52	87.01
16	74.02	96.64	71.00	87.81
17	74.00	96.60	81.67	88.58
18	73.98	96.52	87.20	88.92
19	82.70	59.50	50.48	65.96
20	82.60	59.65	58.58	66.62
21	82.22	59.73	67.42	67.19
22	82.13	78.76	61.58	78.54
23	82.45	78.70	73.13	79.44
24	82.37	78.49	76.71	79.55
25	88.65	59.84	53.31	68.25
26	88.78	59.62	61.54	68.73
27	89.02	59.57	66.89	69.12
28	88.93	78.77	63.53	80.77
29	88.63	78.43	74.63	81.29
30	89.15	78.71	78.18	81.88
31	92.18	59.74	54.35	69.31
32	92.50	59.64	64.35	70.07
33	92.46	59.66	65.60	70.16
34	92.43	78.63	65.76	81.95
35	92.51	78.54	76.34	82.66
36	92.00	78.67	79.47	82.83
37	95.01	59.68	56.76	70.33
38	94.58	59.63	67.23	70.93
39	94.88	59.65	66.68	71.03
40	94.77	78.71	65.24	82.68
41	94.84	78.64	78.09	83.57
42	95.19	78.54	81.96	83.90
43	82.42	87.68	65.66	84.45
44	82.22	87.77	78.91	85.41
45	82.09	87.84	82.72	85.70

Table 3. System Performances of Four Feasible Schemes

Scheme	Staffing Cost ($)	Service Level (%)				Average Waiting Time (min)			Staffing Utilization Rate (%)		
		Type I	Type II	Type III	Overall System	Type I	Type II	Type III	Type I	Type II	Type III
4	1448	59.57	93.70	66.13	81.16	0.39	0.07	0.19	94.71	90.41	74.98
10	1445	74.45	87.88	66.34	82.16	0.24	0.12	0.17	92.93	92.66	73.76
28	1454	88.93	78.77	63.53	80.77	0.10	0.20	0.19	89.13	94.61	74.69
43	1457	82.42	87.68	65.66	84.45	0.17	0.11	0.21	91.39	92.50	75.28