SIMULATION FOR IT SERVICE DESK IMPROVEMENT

PETER BOBER

1 INTRODUCTION

The IT service desk provides a single point of contact for all users of IT service provider. Specific service desk responsibilities include logging, escalating to higher level and closing incidents (an unplanned interruption to an IT service), logging requests of a different kind, answering questions, keeping users informed of the status of services, and managing the lifecycle of incidents and requests (itSMF UK, 2012). The predecessor of service desk is telephone call centre. Comprehensive review of call centre can be found in (Gans, Koole and Mandelbaum, 2003). Service desk managers are expected to deliver high quality of service while operating costs are maintaining low. If we skip intuitive way of doing things, which is a privilege of visionaries, there are several approaches to achieve previous statement.

Practical approach uses best practises and widely accepted standards to organise all resources and processes to run service desk. Mathematical approach is used when quantitative characteristics of the service desk should be calculated, or optimisation is needed. Discrete event simulation takes place when mathematical models are too complex or restricted to special cases.

Service desk is a key function of Service operation as defined in the Information Technology Infrastructure Library (ITIL). According (itSMF UK, 2012) "ITIL is a public framework that describes best practice in IT service management. It provides a framework for the governance of IT, and the management and control of IT services." One part of an IT service life cycle defined in ITIL is Continual service improvement which is seven-step process based on Deming PDCA cycle.

The work of Jäntti and Kalliokoski (2010) may be mentioned as an example of a practical way of improvement. They identified challenges by interviewing service desk staff with carefully chosen questions. Challenges were related to incident classification, automation of incident and order processes, quality of staff instruction documents, and escalation of the incident to the higher level.

Usually the time come when practical approach based on expert guess is insufficient for further improvement of service quality, and more accurate methods are required. Then the mathematical approach is used to solve particular quantitative problem of the service desk operation.

Mathematical approach requires stochastic models for operational parameters e.g. arrival process and service duration. The queuing theory is used to describe the service desk. Solution of such queuing system can be an analytical or numerical and optimal solution for chosen criterion can be evaluated. Comprehensive overview of this topic can be found in the work of Koole and Mandelbaum (2002) and Gans, Koole and Mandelbaum (2003).

To illustrate particular issues to be solved, the work of Dudin and Dudina (2012) is mentioned. They build two-phase queueing system with Markovian arrival flow and try to find optimum staff number with regard to costs for given quality level. They obtain an existence condition for a stationary regime, algorithm for computing stationary probabilities and basic performance characteristics of the presented system. Optimisation problem for the system operation was solved numerically.

Bhulai et al. (2012) provided a stochastic approximation algorithm. This algorithm allows to find the optimal balanced policy to allocate tasks to staff members. The system was described by Markov chain and numerically solved by means of Monte Carlo simulation. Mathematical models are also used for forecasting workload for design of a structure of call centre, calculate staff requirements, and create staff schedules (NAQC, 2010).

Let us move to another topic regarding service desk and call centres efficiency which is individual skills and corporate knowledge base. The S-shaped learning curve is most apparent when someone learns a highly complex task (Dewey, 2014) and the influence of replacement of a staff member can be investigated.

Gonzalez et al. (2001) proposed hypothesis that the knowledge management system will improve the operational performance of call centre. The same hypothesis can be assumed for a service desk. A knowledge management system by gathering organization knowledge would allow each staff member to leverage the organization's knowledge and consequently improve overall service and performance of the service desk. Performance can be improved by knowledge reuse and knowledge sharing implemented in the knowledge base system.

In further works Gonzalez, Giachetti and Ramirez (2005) proposed a new approach, called a knowledge management-centric help desk. The knowledge management system is designed to be incorporated into the daily operation of the help desk in order to ensure high utilization of the knowledge stores. This concept is supported by findings of Jäntti and Kalliokoski (2010), namely by identified challenge "10) The content of some instruction documents has not been updated" that indicates the inconsistency in corporate documentation and emphasise perceived importance of well organised and accurate documentation.

It must be said that not every manager can handle complex mathematical methods or queuing models for optimisations of service desk. Discrete-event simulation (DES) can be more obvious and clear for him. DES can be used in two distinct manners. First, as a single-simulation study to find improvements of the system. Second, as a tool for everyday managers work. The simulation model resulted from a single-simulation study can be reused in everyday decision process of service desk managers.

Single-simulation study approach was used in the work of Akhtar and Latif (2010). In this study, detailed overview of call centre functions and their operations was given, and the effects of calls routing and prioritizing to specific agents with multiple skills in an inbound call centre was discussed and modelled.

Simulation tool for managers was published by Sencer and Ozel (2013). In their study, a simulation-based decision support system was developed that runs on real-time data instead of queuing models that are based on Erlang C calculations, which is oversimplification in some cases. The graphical user interfaces were designed to increase the usability, functionality and effectiveness of the decision support system in accordance with the man–machine interaction consideration.

Bartsch, Mevius and Oberweis (2012), presented another simulation tool based on discrete-event simulation in order to do a priori estimations of workloads within a skills-based IT service desk. Service managers can perform tool-supported capability planning, analysis, and optimization during design time based on the potential impact of workloads to agreed service levels.

This article promotes using discrete-event simulation to define quantitative goals for the service desk improvement. The results of the simulation told managers where they could go width their expectations.

2   METHODOLOGY

This article provides a methodology which uses simulation to help IT service management to make a decision and to plan service strategy. The procedure is illustrated by case examples which clarify application of simulation and provide insight to the service desk operation.

There are several techniques to carry on a simulation study. We distinguish following ones (Kelton, Sadowski, and Sturrock, 2004):

- *What if analysis* is a powerful tool for improvement. It evaluates how strategic, tactical or operational changes may influence the business.

- *System operation analysis* provides insight to the system operation, which helps to define possible changes.

- *Optimisation* seeks system variable values that lead to an optimal value of the selected criteria under existing restrictions.

Procedure for using simulation consists of following steps:

1. Formulate question to be answered – clearly define problem to be solved, choose the technique of using simulation.

2. Build a model which can answer the questions – design, build and debug the model.

3. Plan experiments – define system variables, their values and responses which should be logged.

4. Run experiments – carry on experiments according the plan. Record all responses.

5. Analyse results, find answers – make a report from experiments results.

Sometimes steps should be repeated as new circumstances from the simulation experiments arises.

Following article sections contains case example of the service desk. The model was designed in accordance with existing service desk. The goal of the simulation is:

- to study the influence of staff number to the number of response over time (time exceedance),

- to study influence of staff skill level changing during learning phase to number of time exceedance and average service duration.

Simulation model considers learning ability of service desk agents and growth of knowledge database of the company. Model shows how the learning curve influences the time development of service desk quality and efficiency.

Results of the analysis can lead to change of staff number, to improve training scheme or to plan changes in skill requirements for different service desk level staff.

3   MODEL OF A SERVICE DESK

The service desk logs and manages user questions, request and incidents (service malfunction). It acts as a single point of contact for all other service operation processes and activities (itSMF *UK*, 2012).

Specific service desk responsibilities include:

- Logging requests, incidents and answering questions.

- Managing the lifecycle of incidents and requests, escalating as appropriate and closing them when the user is satisfied.

- Keeping users informed of the status of services, incidents and requests.

3.1 Conceptual Model of a Service Desk

Modelled service desk consist of three levels or layers of service. First level receives incoming calls, e-mails, and monitoring tool alerts and creates documented and traceable logs – tickets. Tickets are sorted to the groups and processed as stated in Table 1. Priority 1 tickets are escalated directly to Level 3 because of customer Service Level Agreement (SLA). Figure 1 shows the conceptual model of service desk. Queues at Level 1 and 3 have FIFO queuing discipline. Queue at Level 2 uses ticket priority queuing discipline according SLA.

Table 1 – Ticket type operation and probability

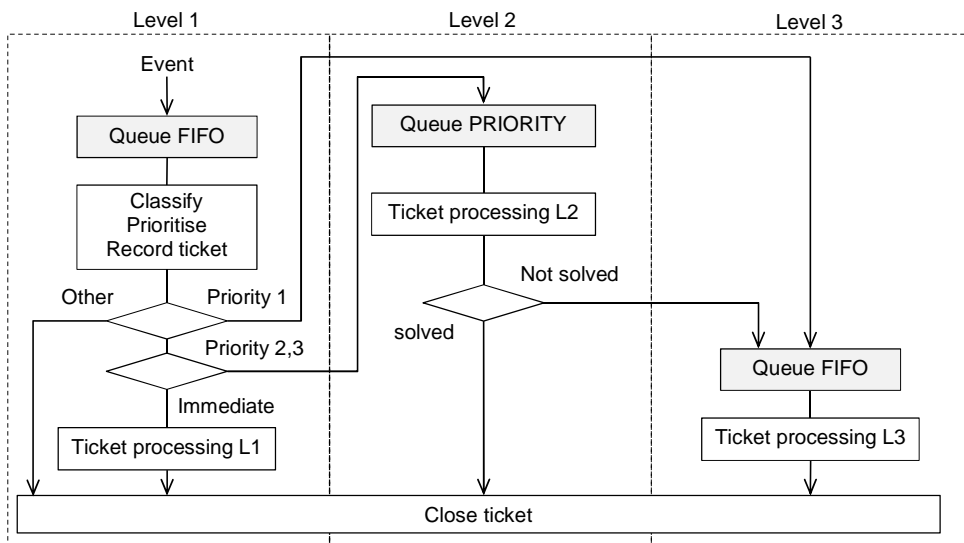| Ticket type	Operation	Probability
Priority 1	escalate directly to level 3	0.08
Priority 2	escalate to level 2, high priority	0.5
Priority 3	escalate to level 2, low priority	0.3
Immediate resolution	serve immediately	0.1
other unit	ticket is not handled at this service centre and is moved to another department	0.02

Figure 1 – Conceptual model of a service desk

3.2 Quantitative Data

Quantitative data were obtained from existing ticketing system and expert's quest of staff members. Interarrival time is seven minutes in average and is modelled by exponential distribution. Processing times was estimated by three-point estimation technique, and they are from a triangular distribution written in Table 2. We must agree with Sencer and Ozel (2013), and Gans, Koole and

Mandelbaum (2003, p. 127) that probability distribution must be critically chosen from collected data, no automatic assumption can be made. But the case described in this article analyses steady state behaviour of service desk model and distributions are stationary.

Table 2 – Service process time parameters and escalation probability

Ticket type	Process time [minutes] Triangular distribution: min, mode, max, calculated average			Escalation to Level 3 probability
	Level 1	Level 2	Level 3	
Priority 1	15,25,45; 28.33	20,300,580; 300	20,300,580; 300	1
Priority 2	15,25,45; 28.33	30,120,210; 120	30,120,210; 120	0.1
Priority 3	15,25,45; 28.33	20,60,100; 60	20,60,100; 60	0.1
Immediate resolution	5,10,15; 10	-	-	-
other unit	4,5,6; 5	-	-	-

3.3 Capacity Calculation

The staff capacity is calculated from already collected data, and results are in Table 3. It uses simple algebraic equations with average values from Table 2.

The quality level of service required an "appropriate" utilization rate. Utilization rates above 80% result more time exceedance (number of over time) as agreed in SLA. Rates below 60% indicate an overstaffed operation and an underutilized workforce (Giva, Inc., 2014). Utilisation for lover staff number was chosen lover (0.75) because of bigger influence on the time exceedance.

Table 3 – Staff capacity calculation (average time between arrivals is 7 minutes)

Ticket type	Probability	Arrivals per hour	Level 1		Level 2		Escalation to Level 3 probability	Level 3	
			Resource ability [tickets per hour]	Required resources	Resource ability [tickets per hour]	Required resources		Resource ability [tickets per hour]	Required resources
Priority 1	0.08	0.686	2.118	0.324	-	0.000	1.00	0.2	3.428
Priority 2	0.50	4.286	2.118	2.024	0.5	8.571	0.10	0.5	0.857
Priority 3	0.30	2.571	2.118	1.214	1	2.571	0.10	1	0.257
Immediate	0.10	0.857	6.000	0.143	-	0.000	-	-	-
other unit	0.02	0.171	12.000	0.014	-	0.000	-	-	-
Sum				3.719		11.143			4.543
Utilisation				0.75		0.8			0.75
Calculated Capacity				4.96		13.93			6.06
Capacity (staff number)				5		14			6

3.4 Basic Metrics

From the customer point of view, there are two main quantitative values which characterise service quality level as agreed in SLA. First is the response time: the time in which it is necessary to respond. Second is the total service time: the total time from the beginning to the end of ticket processing. From the provider point of view, the total service desk costs would be interesting. The first approximation of costs is staff number on different service desk levels.

Table 4 – Required response time and Total service time for different ticket types

Ticket type	Required in SLA	
	Response time [minutes]	Total service time [hours]
Priority 1		12
Priority 2		24
Priority 3	30	48
Immediate resolution		ASAP
other unit		ASAP

3.5 Model Building

Simulation model was built in the Arena simulation software from Rockwell Automation. Influence of staff number and staff skill level should be investigated, therefore corresponding model variables must be assignable (so-called Controls in Arena language). These are: resource capacity for Level 1-3 and processing times on each level. Figure 2 shows a part of the model.
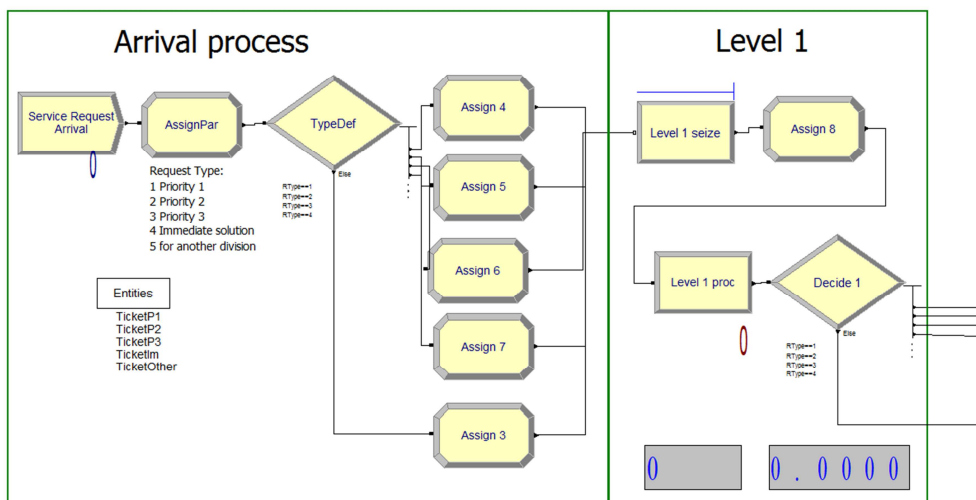

Figure 2 – First two parts of the simulation model in Arena software

Many model characteristics are collected automatically in Arena statistics module, however not everything. Number of response time exceedances was calculated by dedicated blocks. Model was validated against expected capacity

calculations. Resource utilisation on Levels 1-3 were 0.74, 0.79 and 0,75 which corresponds to given values 0.75, 0.8 and 0.75 in Table 3.

4 USING A MODEL

4.1 Staff Capacity Planning

First question to answer in this study was how the staff number influences the service desk response time. The experiment was planned as four simulations consisting of ten runs for different values of Level 1 staff number. The resulting values of a number of response time exceedance were averaged, and confidence interval was checked for validity. Figure 3 shows the results.
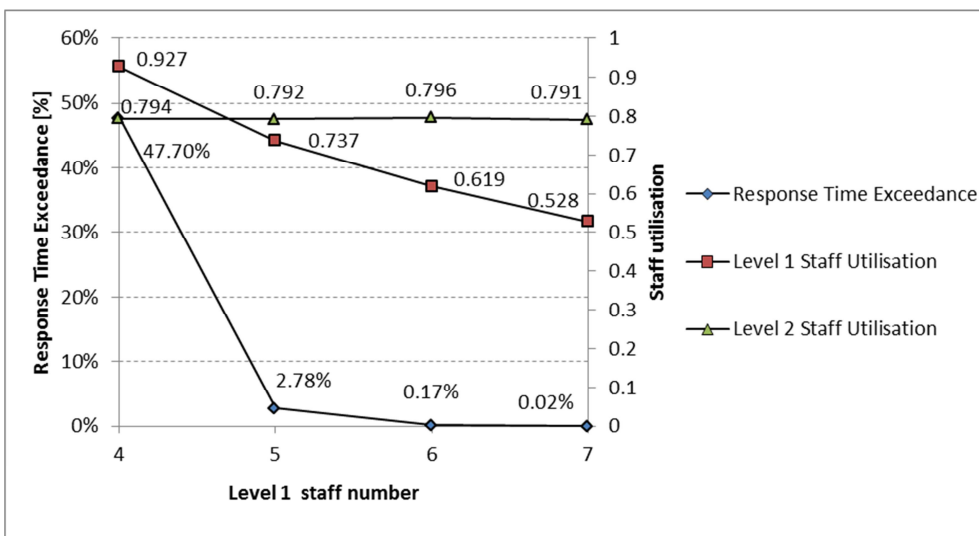
Figure 3 – Response time exceedance and staff utilisation vs. staff number

The finding from experiment is: staff number of 5 is needed for an acceptable percentage of response time exceedance. There are several ways how the staff number can be reduced while maintaining a low exceedance. Then we can ask another question: what happened if staff at Level 2 is helping first level?

To answer this question the model must be modified to reroute requests from Level 1 to Level 2 if waiting time in Level 1 queue is long. The results are on Figure 4 and comparison of first two experiments is on Figure 5. Response time exceedance drops down considerable at four staff members on Level 1. It was paid by slightly rising of Level 2 staff utilisation from 0.794 to 0.814. Such more flexible organisation of staff work can bring remarkable savings. And this is a challenge for deeper analysis.
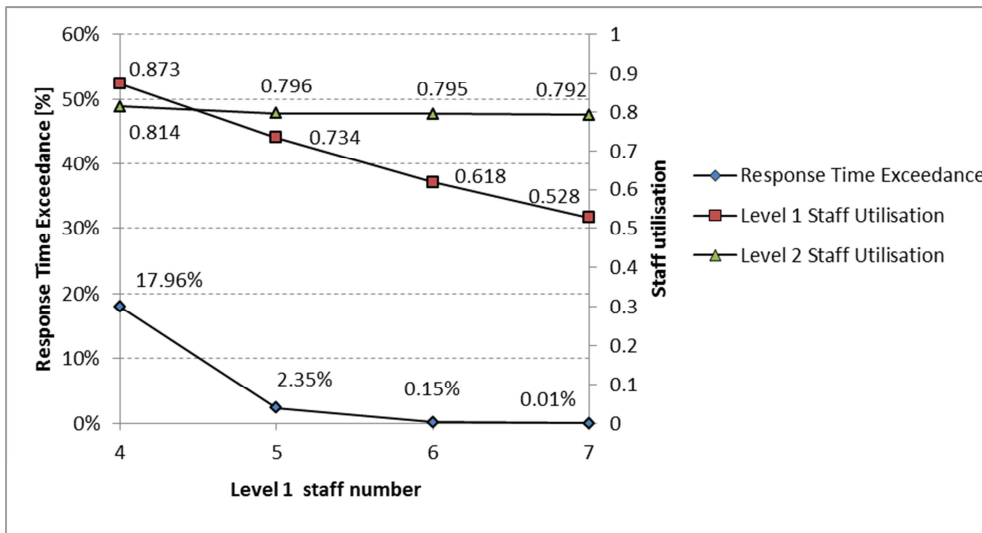
Figure 4 – Response time exceedance and staff utilisation vs. staff number for requests rerouting to Level 2
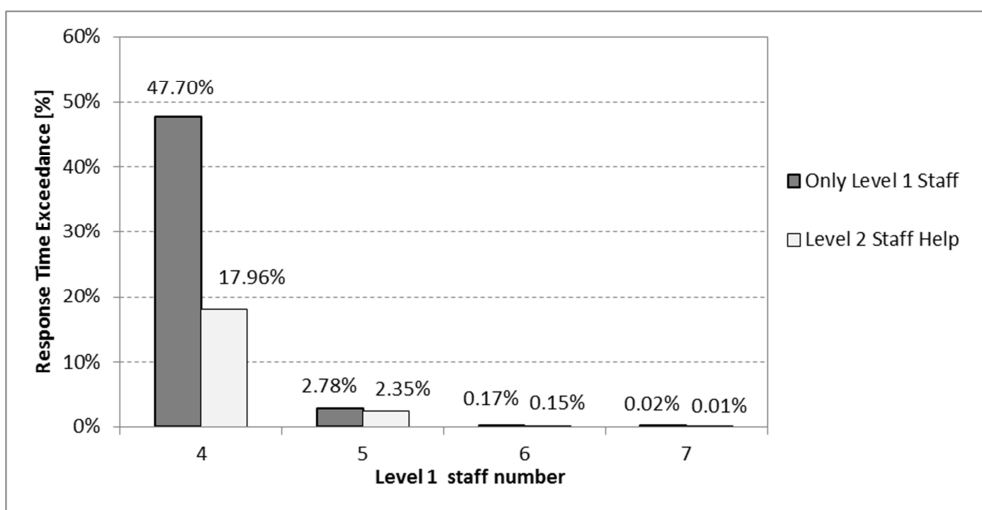

Figure 5 – Comparison of response time exceedance

4.2 Staff Skills and Knowledge Challenges

This part of the simulation study is focused on showing the influence of staff skills to exceedance time. The assumption is that higher level of skills will lover the operating times. Furthermore, the investment to corporate knowledge base can be evaluated, or quantitative target of time shortening can be defined.

In the experiment, the operating times was shortened by coefficient which is following S-shaped learning curve. Figure 6 shows response time exceedance

during learning time five weeks. The S-shaped learning curve represents learning process in which the operational time was shortening by 8%. The staff number at Level 1 is four.
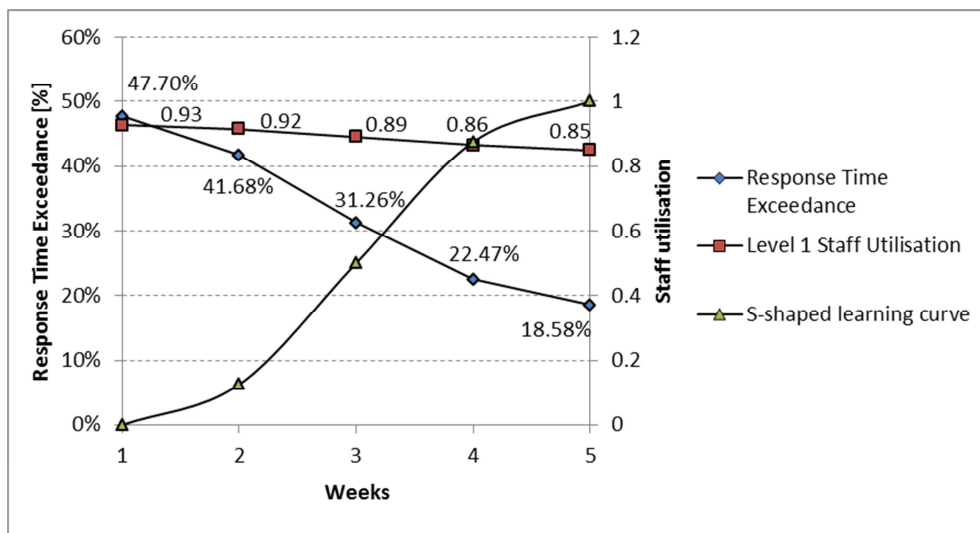

Figure 6 – Response time exceedance reduction by shortening the operational time by 8 % (staff number at Level 1 is four)

The Figure 6 clearly depict that operational time shortening greatly influences reduction of response time exceedance. It means that investments into training or a corporate knowledge base are reasonable in this case.

5   CONCLUSION

This article promotes using discrete-event simulation to define quantitative goals for the service desk improvement. Suggested procedure is explained on service desk model which is built according real one. This article is showing what type of data are needed to build a model and how to plan and evaluate experiments. Performed experiments have shown that the response time exceedance is greatly influenced by workload lightening or operational time shortening. The results of the simulation told the managers where they could go width their expectations in the service desk improvement. Carefully designed experiments can give unexpected insight into service desk functioning.

ACKNOWLEDGEMENT

REFERENCES

Akhtar, S. and Latif, M., 2010. Exploiting Simulation for Call Centre Optimization. In: *Proceedings of the World Congress on Engineering 2010*, Vol III, WCE 2010, June 30 - July 2, 2010, London, U.K

Bartsch, C., Mevius, M. and Oberweis, A., 2012. Simulation environment for IT service support processes: Supporting service providers in estimating service levels for incident management. In: *2nd International Conference on Information, Process, and Knowledge Management, eKNOW 2010*, February 10-16, 2010 - St. Maarten, Netherlands, pp. 23-31, DOI: 10.1109/eKNOW.2010.10.

Bhulai, S., Farenhorst-Yuan, T., Heidergott, B. and van der Laan, D., 2012. Optimal balanced control for call centers. *Annals of Operations Research*, 201(1), pp. 39-62

Dewey, R. A., 2014. *Psychology: An Introduction*. [online] Available at: <http://www.intropsych.com/index.html>, (Chapter 07, Part 3, The Learning Curve), [Accessed 5 June 2014].

Dudin, S. A. and Dudina, O. S., 2013. Help desk center operating model as a two-phase queueing system. *Problems of Information Transmission*, 49(1), pp. 58-72.

Gans, N., Koole, G. and Mandelbaum, A., 2003. Telephone Call Centers: Tutorial, Review, and Research Prospects. *Manufacturing & Service Operations Management*. 5(2), pp. 79-141.

Giva, Inc., 2014. *Help Desk Staffing Models*. [online] Available at: <http://www.givainc.com/wp/help-desk-staffing-models-simple-analysis-save-money.cfm?c=wp_hd> [Accessed 4 June 2014].

Gonzalez, L. M., Giachetti, R. E. and Ramirez, G., 2005. Knowledge management-centric help desk: specification and performance evaluation. *Decision Support Systems*. 40(2), pp. 389–405, DOI: 10.1016/j.dss.2004.04.013

Gonzalez, L. M., Gonzalez, I., Giachetti, R. and Ramirez, G., 2001. Analysis of Applying Knowledge Management in an Information Technology Call Center. In: *Proceedings of the Twelfth Annual Conference of the Production and Operations Management Society*, POM-2001, March 30 – April 2, 2001, Orlando FL.

itSMF *UK*, 2012. *An Introductory Overview of ITIL® 2011*. [online] London: TSO. Available at:
http://www.axelos.com/gempdf/itSMF_An_Introductory_Overview_of_ITIL_V3
.pdf/> [Accessed 4 June 2014].

Jäntti, M. and Kalliokoski, J., 2010. Identifying knowledge management challenges in a service desk: A case study. *2nd International Conference on Information, Process, and Knowledge Management, eKNOW 2010*, February 10-

16, 2010 - St. Maarten, Netherlands, pp. 100-105. DOI: 10.1109/eKNOW.2010.20.

Kelton, W. D., Sadowski, R. P. and Sturrock, D. T., 2004. *Simulation with Arena*. 3th edition, NewYork: McGraw-Hill.

Koole, G. and Mandelbaum, A., 2002. Queueing models of call centers: An introduction. *Annals of Operations Research*, 113(1-4), pp. 41-59.

NAQC, 2010. *Fundamentals of Call Center Staffing and Technologies. Quality Improvement Initiative* (Reynolds, P.). Phoenix, AZ.

Sencer, A. and Ozel, B. B., 2013. A simulation-based decision support system for workforce management in call centers. *Simulation*, 89(4), pp. 481-497.

ABOUT THE AUTHOR

Peter Bober, PhD. is assistant professor at the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia, e-mail: peter.bober@tuke.sk