Operations Management Education Review 18: 83-100.

A Monte Carlo Demonstration of the Implications of Workload Variability in Capacity Planning

Timothy S. Vaughan
University of Wisconsin – Eau Claire, USA

Abstract. This paper addresses the implications of workload variability in capacity planning, and the relationships between nominal capacity, effective capacity, and workload. An Excel-based Monte Carlo model provides a graphical depiction of the implications of workload varying relative to some given capacity. A guided exercise demonstrates that workload variability necessitates some amount of "buffer" or "capacity cushion" relative to average workload. In the face of variable terminology that may impair students understanding of the issue, the demonstration also helps to clarify that the buffer of concern exists for reasons other than possible sources of planned or unplanned process downtime.

Keywords: capacity planning, utilization, efficiency, Monte Carlo simulation.

1. Introduction

Capacity planning and scheduling are inter-related activities that occupy a central role in operations and supply chain management. Scheduling generally represents decisions regarding how to best arrange a given workload, in order to satisfy objectives within the constraints of existing capacity. Scheduling problems may however raise the issue of capacity adjustments, particularly those related to short-term, flexible capacity alternatives.

Capacity planning involves decisions regarding the quantity of various resources that should be in place, in anticipation of future workloads. Short-term planning might address adjustment of flexible capacity alternatives, relative to projected workloads as computed from some firm production schedule or short-term demand forecast. Longer term capacity planning addresses decisions regarding less flexible resources that are not economically varied in the short term. Longer term capacity planning typically addresses a greater degree of uncertainty relative to short term.

As a separate concept, both short-term and long-term capacity planning must address predictable and/or random variability in workload. In the face of such variation, capacity planning and scheduling involves the options of (a)

varying capacity in response to variation in workload and/or (b) leveling the workload to fit within some given capacity. Under option (b), workload leveling may take the form of pulling work into an earlier period (if variations in workload can be predicted in advance), and/or allowing the work to backlog into a later period. In either case, the degree of workload leveling required depends on the actual capacity provided *vis-à-vis* the variable workload pattern. As such, setting capacity levels requires attention to the fundamental trade-off between resource utilization vs. system responsiveness in the face of variable workloads.

This paper is concerned with developing students' understanding of this fundamental issue. The following section presents a simple conceptual model designed to clarify and distinguish between separate concepts within the capacity planning discussion. A discrete-time Monte Carlo simulation is then introduced, supporting a guided in-class exercise that provides a compelling depiction of the implications of workload variability when setting capacity levels. The demonstration clarifies that it is workload as a percentage of *effective* capacity, (as opposed to workload as a percentage of *nominal* capacity) that is of primary interest when deciding on capacity levels in the face of a variable future workload. The simulation model provides a readily accessible visualization of the effects of workload variability, in an introductory course that does not allocate time to developing student understanding of queuing models or more complex discrete-event simulation.

2. A Capacity Planning Framework

Capacity planning requires a clear understanding of two related but very distinct concepts, namely workload vs. capacity. *Capacity*, in short, is some measure of how much work could be done within a specific time period. *Workload*, in contrast, is a measure of how much work has been done or will be done within the same time period. Capacity planning thus requires measures that support meaningful comparison between workload and capacity.

Nominal vs. Effective Capacity

In turn we have two different measures of capacity that typically enter the capacity planning discussion. Most capacity discussions begin, appropriately, with some discussion of the difference between what might be termed *nominal* or theoretical capacity, vs. some reduced measure referred to as *effective* capacity. Although the specifics of the definitions vary across textbooks, nominal capacity generally refers to some output rate that could be realized under ideal conditions over some short period of time. As a simple example, a production process capable of producing one unit every 2 minutes would have