

Modeling the Worker

David C. Juran

*Graduate School of Business
Uris Hall, Columbia University
New York, NY 10027*

*Stern School of Business
Tisch Hall, New York University
New York, NY 10012*

Lee W. Schruben

*Department of Industrial Engineering and Operations Research
Etcheverry Hall, University of California
Berkeley, CA 94720-1777*

Revised 12/1/2000

Submitted to Journal of Operations Management

Modeling the Worker

Abstract

This paper deals with worker performance variation in unpaced systems, in which human factors constitute a significant set of productivity constraints. Examples include highly technical industries (e.g. semiconductor manufacturing) and service industries (e.g. banking) that rely increasingly on fewer but more highly skilled workers. In these systems, productivity is dependent on the availability and organization of vital human resources, and it is reasonable to expect that performance models of these systems will need to contain better representations of worker behavior than are available using conventional modeling methods.

This paper examines one issue of relevance to modeling human workers: the problem of replacing random variables (estimates of parameters from empirical data) with constants for the purposes of simulation experiments. We offer here one possible approach by which one can account for the inherent randomness in parameter estimators rather than ignoring it. We use the specific example of worker task time parameters to illustrate the more general problem of addressing the variability of input parameter estimators.

We also explore the tradeoffs in including information about the demographics and personalities of workers in system performance simulation models. This paper reports the results of a series of actual and simulated experiments in which personality and demographic data are used in different ways to model the performance of a work group. Significant differences are found in predicted system performance and model validity depending on the methodology used for modeling workers. These results have practical implication for the managerial processes of recruiting and selecting individual workers, as well as organizing teams of workers and assigning them to productive tasks.

Key words: Human resources/ OM interface, Interdisciplinary, Job design, Empirical research, Measurement and methodology, Service Operations, Computer simulations, Staffing, Work measurement, Productivity, Personnel and shift scheduling.

1. Introduction and Background

Highly technical industries such as semiconductor manufacturing and service industries such as banking continue to rely increasingly on fewer but more highly skilled workers. As this trend continues, productivity will become increasingly dependent on the availability and organization of vital human resources, and it is reasonable to expect that performance models of these systems will need to contain better representations of worker behavior. When locating a new production facility, expanding current production, or moving into a new service area, the demographics of the local labor force may be an important consideration. As the politicians from high-cost regions would have us believe, the prevailing local labor rates may not tell the whole story. For example, the availability of educated, tech-savvy, motivated, and honest workers might be more important than the average labor cost in a given region. However, unless operations models are able to account for these human factors, we will not be able to quantify how much these characteristics are actually worth to an organization's bottom line.

Researchers have attempted to address this modeling issue through a number of reasonable approaches, involving various combinations of psychological and operations research disciplines. This crossover work between engineering and psychology has taken several forms. Some organizational science work examines the effects of psychological independent variables on other psychological variables, such as work group processes (see, for example, Barry and Stewart, 1997). Other researchers in the field of operations have examined the effects of structural system variables on operational outcomes such as throughput (e.g. Conway et al., 1988).

Increasingly, interesting discoveries have been made through collaboration across disciplines. In Doerr et al. (1996) and Schultz et al. (1999) effects are established between operational system elements and psychological outcomes. Conversely, there has been some crossover work that examines how psychological variables affect operational outcomes (Juran 1997; Lin and Chu, 1998). In an even more complex approach, Schultz et al. (1998) explores how an operations variable (low inventory) works through a psychological variable (group norms) to produce an unexpected operational effect (processing time variability).

This paper focuses on operational modeling aspects of the effects of psychological variables on productivity outcomes. The existence of an effect, once established, provides some rationale for including that effect in an operational model, but it does not provide a modeling methodology for doing so. This paper is aimed at providing some of that methodology.

1.1 Parameters and Estimators

All good stochastic modeling textbooks caution their readers to be wary of replacing a random variable with a constant. For example, replacing an input random variable in a production system model with an estimate of its expected value can cause a gross underestimation in the variability of the actual production system.

Despite this well-known principle, it is universal modeling practice to “determine” (Law and Kelton, 1991; p. 395) the values of parameters for probability distributions used in simulations and treat these random observations as known constants for the purposes of simulation experiments. Kelton (1984) and Banks and Carson (1984), for example, provide quite similar descriptions of this practice, consisting of the following basic steps:

1. Collect data (actual observations of the random variable to be modeled),
2. Identify a distribution type that seems to fit the data (e.g. exponential, beta, etc.),
3. Estimate the parameters (e.g. λ for exponential), and,
4. Conduct Goodness of Fit Tests (e.g. Chi-square, Kolmogorov-Smirnoff)

Pegden, Shannon, and Sadowski (1995) note one of the difficulties with this methodology: “It should be noted that goodness-of-fit tests such as these generally have a low probability of rejecting an incorrect fit. As a result, these tests often repeatedly fail to reject a fit when the same set of data is tested against several different distributions. The fact that the test does not reject the fit for a distribution should not be taken as strong evidence that the selected distribution is a good fit.” Through the use of distribution fitting software (UniFit, BestFit, @Risk, or Arena), one can extend this procedure, and identify the “best” among several theoretical distributions.

Kelton (1984) provides some details of this methodology with respect to parameter estimation, a confidence interval method for sensitivity analysis. While this is in fact a useful method for sensitivity analysis regarding the value of an input parameter, it doesn't solve a more basic problem; the methodology described above replaces random variables (in this case estimates of input parameters) with constants (in this case maximum-likelihood estimators or other statistical estimates of the unknown parameters).

Kelton, Sadowski, and Sadowski (1998) identify the fundamental problem we attempt to address in this paper: “Since we're estimating a parameter by just a single number (rather than an interval), this is called point estimation. While point estimates on their own frankly aren't worth much (since you don't know how close or stable or generally good they are), they're a start and can have some properties worth mentioning.” This issue is also acknowledged in Schmeiser (1999), which discusses the differences between Helton's (1996) stochastic and subjective uncertainty, but which also concludes that “the state of the art is far from allowing novice practitioners to build complex input models in the way that they can build complex logical models with today's commercial software”.

We offer here one possible approach by which one can account for the randomness in the parameter estimators rather than ignoring it. We use the specific example of worker task time parameters to illustrate the more general problem of addressing the variability of input parameter estimators. In this paper we also explore the tradeoffs in including information about the demographics and personalities of workers in system performance simulation models. This represents one answer to the problem of subjective uncertainty, but is not the only answer. It may be possible to improve the predictive validity of operations models by better representing the differences among individual workers and their effects on system performance. If this is the case, we will need to learn how to model machines and people with “equal fidelity” (Kempf, 1996).

1.2 Approaches to Recognizing Randomness

We start with a view of simulation modeling that considers the output of a simulation experiment to be a function of two basic elements: (a) input parameters, here represented by the vector X , and (b) the

simulation model itself, here represented by S . On the most basic level, then, the results of an experiment, Y , can be represented:

$$Y = S(X)$$

We further assume that X has some known distribution F , with an unknown parameter set P . Using this distribution (as specified by P), researchers can employ some method of randomization U — typically a pseudo-random number stream — to actually perform experiments.

$$Y = S[F(P, U)]$$

The focus of this paper is this parameter set P , which is traditionally created using the methods described in Section 1.1 above. An empirical sample (of size N) of the target system's performance can be observed, and the elements of P estimated on the basis of some set of statistical estimators G :

$$P = G(X_1, X_2, X_2, \dots, X_N)$$

We propose several extensions to this basic method, some of which offer significant advantages with respect to the validity of the simulated system.

Random Parameters. Note that, although all of the distribution parameters in the conventional method are estimated from sample data (and therefore subject to sampling error), their values are treated as correct and known constants. Instead, we propose modeling the input parameters as random variables, which themselves have parameters. In Section 2.2 below we illustrate this method with a processing time example. The processing times used in the simulation are generated from a distribution with randomly generated parameters based on empirical observation.

Randomly Selected Teams. In systems with human entities, one source of variability is variability across teams of workers. In Section 2.3 below we simulate selecting a new team from a “labor pool” for each replication of our experiment. The basis for assigning vectors of input parameters to simulated teams comes from the observed performance of different teams in an empirical setting.

Using Demographic Information. While the previous approach acknowledges variability across samples of worker teams, it ignores demographic information, which can help identify and explain the

sources of variability across workers. We suggest a further extension to the methodology to make use of information about the attributes of individual entities (in our example, personality and demographic data about specific workers).

In the course of empirical research aimed at estimating the parameters of input variables, it is also possible to collect other information about the entities being studied. If sufficient reason exists to consider this set of other information (represented by the vector V) to be useful in explaining variability in performance¹, we might consider treating the observed performance of certain entities as a dependent variable in a model such as:

$$X_i = C_1V_1 + C_2V_2 + C_3V_3 + \dots + C_mV_m$$

In which X_i is the i th observation of the entity's performance, V is an m -dimensional vector of attributes that describes the entity, and C is an estimated vector of the effects on X associated with the attributes V . C is estimated using regression analysis, on the basis of empirical observations of actual entities. Using empirical distributions of the attributes V , we can simulate the selection of different entities (teams of workers in our application), and model their performance taking into account the effects of

¹ Meta-analytic work among personality psychologists (Barrick and Mount, 1991; Ones et al., 1993) has made use of the so-called "Big Five" dimensions of personality to establish their predictive validity with respect to job performance. More recently, operations management scholars have conducted behavioral experiments (Doerr et al., 1996; Juran, 1997; Schultz et al., 1998; Schultz et al., 1999), indicating that measurable behavioral factors, including individual differences, are important in explaining variability across teams of workers in production systems. For a more complete discussion of the Big Five, see Goldberg (1990) or Mount and Barrick (1995); "thumbnail" definitions from Costa and McCrae (1992) are as follows:

Neuroticism (N) assesses adjustment (as opposed to emotional instability) and identifies individuals prone to psychological distress, unrealistic ideas, excessive cravings or urges, and poor coping responses. Neuroticism, briefly described, is a general tendency to experience emotional distress such as embarrassment, anger, or disgust.

Extraversion (E) assesses quantity and intensity of interpersonal interaction, activity level, need for stimulation, and capacity for joy. Extraversion is a general tendency to enjoy the company of other people, and to be talkative, assertive, and active.

Openness (O) assesses proactive seeking and appreciation for experience for its own sake as well as tolerance and exploration of the unfamiliar.

Agreeableness (A) assesses the quality of one's interpersonal orientation along a continuum from compassion to antagonism, in thoughts, feelings, and actions.

Conscientiousness (C) assesses the degree of organization, persistence, and motivation in goal-directed behavior, and contrasts the dependable and fastidious with those who are lackadaisical and sloppy. In the literature, Conscientiousness has consistently been positively associated with work performance across many types of jobs.

their individual attributes (personality and demographic characteristics in our example). We illustrate this method in Section 2.4.

Some of the elements of C in our sample can be “determined” (e.g. a worker's age), while others must be estimated (e.g. personality attributes, which can be measured using psychometric instruments such as the example in Section 2.3).

We can then use demographic information to estimate the distributions of the elements of V in our population. The $C_1, C_2, C_3, \dots, C_N$ we observe in our empirical study are a random sample with distributions $R(M)$. (R for “real” - having unknown parameters M .) Of course, we are still doing some degree of estimation when we do this, but we have the advantage of using a much larger sample to estimate M . With this approach we can infer information about the distribution of C and include knowledge of the randomness in the population at large into our model. This allows us to account for variability in our sample and in our parameter estimators, and, finally, include this source of randomness in Y .

Other variations of this method can be used to perform many replications of simulation experiments based on a single entity (a team of actual workers), as in Section 2.5, or based on some hypothetical entity, as in Section 2.6.

The thrust of this method is twofold: First, we account for randomness in the parameters in the simulation model. Second, we attempt to exploit our empirical knowledge of the entity population (labor pool) from which the data were collected to better estimate this randomness, without requiring a larger sample.

Several approaches to modeling individual worker differences are demonstrated below, using these extensions of standard simulation methodology. Six simulation experiments and one laboratory experiment were performed to study different ways of using personality and demographic information about individual workers.

The simulation experiments in this paper demonstrate varying degrees to which information about workers might be included in a simulation model. Some of these modeling approaches require extensive behavioral data and analysis similar to that in Juran (1997). Whether modeling people in greater detail is worth the extra effort ultimately depends on the stakes and risk in the system being studied.

1.3 The Target System

The system we studied is a two-worker, three-machine serial worksharing production cell, in which all three machines are identical. To motivate our experiments, it is useful to regard this cell as part of a larger operation, because randomness of this cell may greatly influence the performance of the entire system. Therefore, we wish to model accurately the actual random behavior of this cell rather than merely to estimate, say, its average performance. This is the cell studied in the behavioral laboratory described in Juran (1997), which has the same fundamental structure as the minimal serial worksharing system in Zavadlav et al. (1996). For more information about such serial worksharing systems, see Bartholdi and Eisenstein (1996), or Bischak (1996).

For the purposes of this paper, we view the behavior of the system in the lab (operated by human workers) as normative, in the sense that we are trying to develop a simulation method that will accurately replicate this system. This behavior is not normative in the sense that it represents the “best” possible configuration of the system, nor is there any 100% assurance that the lab experiment isn’t subject to sampling error².

Using the results of the behavioral experiment as a basis, we perform six simulation experiments, as described in the following sections. Experiment 1 is intended to be representative, in some sense, of conventional modeling practice; all “workers” are assumed to have the same underlying distributions of processing times. In Experiment 2 we apply a Bayesian method, treating processing time parameters as random variables themselves. In Experiment 3, each team has a randomly drawn set of processing time parameters. In Experiment 4, we vary the processing time parameters of the various teams of workers,

² The subjects in the lab experiment were United Steel Workers union members employed at a die-casting plant in Connecticut. Details of that experiment appear in Juran (1997).

explicitly modeling them as dependent variables driven by variation in demographic variables.

Experiment 5 contains many simulation runs in which the two workers do not vary. In Experiment 6, we study the system with two “generic” workers, who are demographically “average” in every way.

Figure 1 illustrates the cell as simulated and as it was set up in the behavioral laboratory. In this cell, two-worker teams process jobs on three machines arranged in a serial line. Jobs proceed through the cell from left to right. The jobs performed in the laboratory were chosen to be typical of those found in high-technical manufacturing such as in semiconductor production and in service operations such as call or order fulfillment centers. The actual tasks involved reading, checking, and entering information on a computer keyboard.

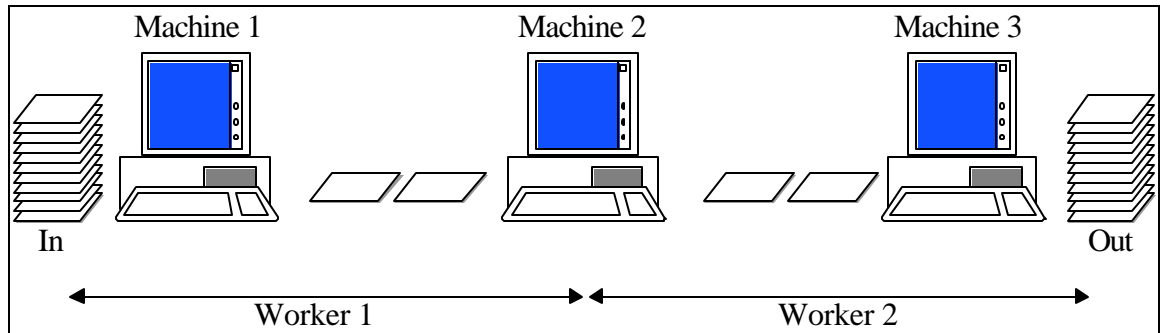


Figure 1: The Laboratory Factory

In order to focus on worker behavior, the laboratory system was balanced in the sense that every job involved the performance of three identical tasks, each on one of the machines. Each task consists of a random set-up time, a random processing time, and a random post-processing or takedown time. Between tasks, a worker also spends a random amount of time moving between machines, referred to here as the move time. The workers are cross-trained and will be assigned machines within the serial line: Worker 1 is assigned to Machines 1 and 2 and Worker 2 is assigned to Machines 2 and 3. The workers operate one of their two assigned machines on the basis of the state of work-in-process inventory, in accordance with pre-determined rules. The workers are paid according to an incentive system designed to promote teamwork (Lawler, 1976).

Occasionally a worker may “bump” his/her partner from an operation, which entails some lost time, called bump time. Bumping is an informal mechanism by which the workers resolve the issue of which of them has precedence at the shared machine (Machine 2); a complete description of bumping behavior for this system appears in Zavadlav et al. (1996). The performance measure of interest is the production cell’s processing rate, measured as the number of jobs per hour. Figure 2 is a histogram of the observed cell performance using actual manufacturing workers under carefully measured conditions, as reported in Juran (1997). For comparison purposes, all histograms in this paper are shown scaled to represent probability densities (their areas are constant).

There were 24 workers involved who were randomly paired into 48 teams. The worker teams were formed according to a Latin-hypercube sampling scheme with each worker being in the upstream position and in the downstream position twice and no two workers being on the same team twice regardless of position. The method of using each worker multiple times, and in both line positions, was employed as a means for helping to exclude the possibility of one form of sampling error. Any differences noted between the two line positions cannot be attributed to having used different pools of workers in the two positions.

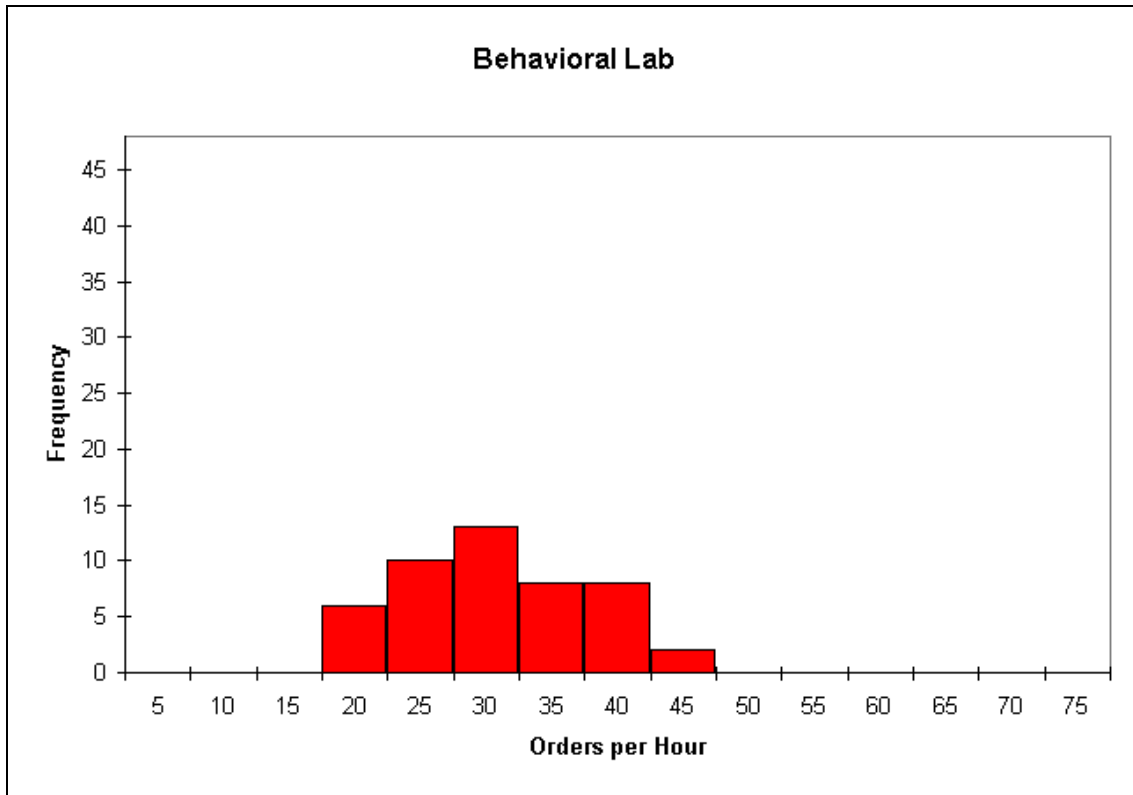


Figure 2 - Histogram of Behavioral Lab Results

1.4 The Simulation

A simulation model representing the production cell was written in C using SIGMA (Schruben, 1992). The modeling approach we used provided an unusual degree of confidence that the logical relationships between system events in the laboratory were being modeled correctly. In fact, the event graph paradigm used here, along with the detailed monitoring of events that is possible in this controlled laboratory, allowed this system to generate its own simulation logic. The method here is not dependent on the use of SIGMA; one reviewer of an earlier version of this paper suggested that one might be able to use “human-in-the-loop” software³, which is an interesting suggestion worth further consideration.

In the early stages of development of the behavioral laboratory, a serious effort was made to create and maintain concurrently a parallel simulation model of the system in the spirit of Gross and Harris (1985, p. 482):

³ See, for example, WinCrew, from Micro Analysis and Design, Boulder, CO 80301.

“A far more formidable aspect of model validation lies in deciding whether the model is an adequate representation of the real world that it is intended to describe. ...If the model is a description of an actual on-going system for which historical data are available, these data can be used as input to the model and the subsequent model output checked with actual history.”

The simulation model was enriched and validated by conducting parallel trace-driven simulations from actual laboratory data with real workers. For each run, data were automatically collected and transformed into corresponding times in the simulation model. These actual inter-event delay times were used to drive the events in the parallel simulation run. The simulation was run until its sequence of events diverged from the sequence of events recorded in the lab. Each time this happened, a new relationship between system events was discovered and added as an edge in the event graph or (rarely) the state or event space of the model was enlarged. The laboratory was designed to be “self-simulating”.

This modeling process led to a number of refinements to the simulated system. Running parallel trace-driven simulations made it clear when the set of precedence relationships between events was incomplete; the workers sometimes acted on the basis of cues not represented in the simulation model. Once uncovered, these new event relationships were defined, labeled, and added to the simulation model until repeated sets of laboratory data could be replicated *exactly* by the simulation model. This process was continued until the parallel simulation consistently produced behavior identical to the laboratory and conceivably could be automated to shadow every laboratory experiment.

With modern automatic event recording mechanisms, such as WIP tracking systems in semiconductor factories or computer tracking systems in call centers, it is conceivable that very accurate simulations of these systems could be generated in this manner. The potential for semi-automatically creating virtual systems is intriguing. Self-simulating systems would have the downside of probably being far too detailed for all but the most demanding fidelity. The general applicability and utility of this method of generating system simulations is currently under study. It seems most promising in systems where workers assert a high degree of control and data collection is highly automated. For the purposes of this paper, the approach provided us with simulation event logic in which we had a great deal of confidence. This allowed us to concentrate on modeling the input processes used to drive the simulation.

1.5 Modeling the Input Processes

In our simulations, we assume that distributions of set-ups, take-downs, item processing times, move times, and bump times may be different for different workers, but were independent of the machines or tasks (which were, in fact, identical). The residual errors in processing times fit a Normal probability distribution quite well. To avoid generating negative times we conditioned the noise to be positive. Variate rejection caused by this conditioning was insignificant, occurring with a probability of less than two percent. We will denote the conditional positive-normal⁴ probability distributions with mean, m , and standard deviation, s , as $PN(m,s)$. As we will see, these PN models provided a very accurate representation of production cell behavior once demographic and personality factors are taken into account.

The simulation requires 10 input variables: Worker 1 Set-Ups, Worker 1 Items, Worker 1 Take-Downs, Worker 1 Moves, Worker 1 Bumps, Worker 2 Set-Ups, Worker 2 Items, Worker 2 Take-Downs, Worker 2 Moves, and Worker 2 Bumps. There are, therefore, twenty input parameters required for each simulation run (a mean and standard deviation for each of ten PN distributions). The experiments in the remainder of this paper focus on different ways of modeling these twenty parameters. Using the data collected in the behavioral lab, we modeled these parameters to different levels of detail in the simulations. Each simulation run was for 10 million time units (the time unit used is 1/60 second). Each experiment consisted of 100 runs.

2. Modeling Worker Populations

In this section we will look at four different approaches to modeling populations of workers. Comparisons of the results show that including information on the distributions of worker demographics and personalities dramatically improves model validity.

2.1 Experiment 1 - Conventional Methodology

⁴We use the term positive-normal here to indicate that negative values were excluded from these distributions.

In this first experiment, we used a methodology to model our production cell that is intended to be representative of the conventional simulation practice as described in Section 1.1. This conventional modeling methodology assumes that all workers have the same underlying distributions of processing times. For example, the distribution of the set-up times on Machine 1 does not change when one worker is substituted for another in the simulation model. While there is variation in processing times, this variation is assumed to be independent of the workers.

The methodology was applied here as follows: (1) teams of workers were selected at random; (2) their performance characteristics were measured; (3) these measurements were used to select probability distributions; (4) the parameters for these distributions were estimated; and (5) the processing times used in the simulation model were randomly generated using these fitted distributions.

Here we assume that all differences between workers are irrelevant to the system's performance. Human variability can be modeled using probability distributions whose underlying parameters are independent of the individual workers who are operating the system. The times used in this simulation experiment were

$$D_{jkl} \sim PN(\hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)$$

D_{jkl} is the random processing time for the random inter-event time $j \in [\textit{worker 1 set-up, worker 1 item, worker 1 take-down, worker 1 move, worker 1 bump, worker 2 set-up, worker 2 item, worker 2 take-down, worker 2 move, worker 2 bump}]$ when worker k is in the upstream position and worker l is in the downstream position. $\hat{\mathbf{m}}_j$ is the sample mean time (estimate of \mathbf{m}_j) for j observed in the behavioral laboratory (including all workers in all runs), and $\hat{\mathbf{S}}_j$ (estimate of \mathbf{s}_j) is the corresponding estimated standard deviation.

In this illustration of our methodology, data are collected in the laboratory from observations of N teams. Each team consists of worker k in the upstream position and worker l in the downstream position. In our case, we used 24 workers in various combinations to comprise $N = 48$ teams (each

worker operated the upstream position twice, and the downstream position twice, all four times with a different partner).

During our discussion we sometimes refer to workers as individuals (e.g. “Upstream Worker k ”) and at other times refer to them as a team (e.g. “Team kl ”). The distinction should be clear from the context. When the positions aren't relevant, kl indexes teams. Therefore, k and l have ranges from 1 to 24, while kl (treated as a single index of the teams) has a range from 1 to 48.

We simulated this production cell 100 times with different random number seeds, using the same input parameters. A histogram of the results appears in Figure 3. This is a very “tight” distribution; using the same scale as Figure 2, all 100 data points fall into a single “bucket” in the histogram. This simulation-generated distribution suggests that the cell performance is much less variable than the real system as observed in the behavioral laboratory.

The result is not surprising; the only source of randomness in the model is the random nature of the individual processing times. Since they all have the same underlying mean, a long simulation run is very likely to result in an estimated population mean that is very close to the true population mean. The point is that this is not a very good reflection of the actual system's behavior; in the lab, the different runs had very different average processing times. This gross underestimation of a single production cell's variability may result in a simulation being wildly optimistic in estimating overall system performance.

2.2 Experiment 2 - Random Parameters

Note that, although all of the distribution parameters are estimated from sample data, the conventional simulation approach assumes that, once they are estimated, their values are all correct and known constants. An error in specifying m_j causes the underlying production rate of the simulation to be different from the system it is meant to represent. An error in s_j for a production cell can effect the performance of the overall system; this is particularly true in a tightly coupled system, such as for JIT production or assembly, in which variation is a major cause of congestion.

To improve our results we will try to account for the fact that we do not actually know the true values of the underlying parameters for the processing time distributions. There are numerous approaches we might apply here, including bootstrapping the original data sample and Bayesian methods. We choose here to use the Bayesian normal model with non-informative or diffuse prior distributions (Gelman, et al., 1995). We ignore the error due to possible variate rejection. With this model, we can include one additional level of uncertainty about the parameter values although we are still assuming that workers are *exchangeable* (in the Bayesian sense). This is done by modeling the processing time parameters as random variables which themselves have normally distributed means and inverse- χ^2 distributed variances. For a particular event delay time, let y_{ij} denote the i^{th} observation, $i = 1, 2, \dots, n_j$. First, the sample mean and variance are computed from the n_j pooled observations of each of the processing times,

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \quad (\text{i})$$

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 \quad (\text{ii})$$

We then can generate a variance parameter, \mathbf{s}_j^2 , from the inverse- χ^2 distribution with $n_j - 1$ degrees of freedom. That is, generate X from a $\chi_{n_j - 1}^2$ distribution and set

$$\mathbf{s}_j^2 = \frac{(n_j - 1)s_j^2}{X}$$

The mean parameter, \mathbf{n}_j , can then be generated from a normal distribution with mean equal to the sample average and variance parameter \mathbf{s}_j^2/n . Finally, the processing time used in the simulation is generated from a (positive) normal distribution with the randomly generated mean, \mathbf{m}_j , and standard deviation, \mathbf{s}_j .

$$D_{jkl} \sim PN(\mathbf{m}_j, \mathbf{s}_j).$$

The results of 100 simulation experiments are shown in Figure 4, where we see that accounting for parameter uncertainty (with a diffuse prior) had little effect due to the large sample sizes involved as long

as the workers are considered identical. We conclude that this approach doesn't add much marginal validity to our simulation results, with respect to the laboratory system. Although the input parameters are treated as random variables, they in fact still are being drawn from common distributions, whose parameters are constant over all 100 runs.

2.2 Experiment 3 - Randomly Selected Teams

A non-Bayesian approach to accounting for the fact that we do not know the "true" parameter values for the delay time distributions has the following motivation. We consider the 48 teams we observed as being a representative sample of worker teams from some larger labor pool. We will simulate selecting a new team from this labor pool for each replication.

For this simulation experiment, processing time j is generated using

$$D_{jkl} \sim PN(\bar{X}_j, s_j)$$

If a different set of teams provided the data from this same labor pool, then, for a particular event delay time, the values of \bar{X}_j and s_j would be different. We will therefore regard \bar{X}_j and s_j as themselves pseudo-normal random variables with distributions fit to the data from the 48 observed teams. For a particular event delay time, let y_{ijkl} denote the i th observation of the inter-event delay j for team kl . First, the sample mean and variance are computed from the data for each team,

$$\bar{y}_{jkl} = \frac{1}{n_{jkl}} \sum_{i=1}^{n_{jkl}} y_{ijkl}$$

$$s_{jkl}^2 = \frac{1}{n_{jkl} - 1} \sum_{i=1}^{n_{jkl}} (y_{ijkl} - \bar{y}_{jkl})^2$$

In the above formula s_{jkl}^2 is the estimate of the variance for an event delay time j for team kl (recall that here kl is regarded as a single subscript, indexing the teams). We compute the average and the variance of its variance

$$\bar{s}_j^2 = \frac{1}{n_j} \sum_{kl=1}^N s_{jkl}^2$$

$$t_j^2 = \frac{1}{n_j - 1} \sum_{kl=1}^N (s_{jkl}^2 - \bar{s}_j^2)^2$$

For each simulation run we simulate selecting a new team kl by sampling \bar{X}_{jkl} from

$$PN\left(\bar{y}_j, \frac{s_j}{\sqrt{n_j}}\right)$$

where \bar{y}_j and s_j^2 are defined as in equations (i) and (ii) on page 16. We sample a new variance, s_{jkl}^2 , from

$$PN(s_j^2, t_j)$$

We ran this simulation 100 times for 10 million time units each, using the same 100 random number streams as in Experiment 1, but with a different randomly drawn set of input parameters for each run. A histogram of these runs appears in Figure 5. This distribution is broader, better reflecting uncertainty as to production rates across teams of workers.

While this approach acknowledges variability across samples of worker teams, it ignores demographic information, which, as noted in Juran (1997), helps us identify and explain the sources of variability across workers. In practice, building a model like the one described here in Experiment 3 requires something like replicated, jackknifed, or batched data to estimate the randomness in the input parameters for simulation modeling. If demographic information can be ignored, the 48 teams used here provide replicated data.

2.3 Experiment 4 - Using Demographic Information

The modeling methodology is extended here to make use of personality and demographic data. Processing time differences between teams of workers are regressed against the personality and demographic attributes of the workers. In practice, this would require data collection and subsequent regression analysis similar to that described in Juran (1997). In that behavioral research project, it was found that individual difference variables explained 80% of the variation in the productivity of serial worksharing teams.

For this experiment we define a team consisting of Worker k in the upstream position and Worker l in the downstream position in terms of a 14-dimensional vector, whose elements are personality and demographic variables⁵:

$$V_{kl} = [N_k, E_k, O_k, A_k, C_k, Age_k, Com_k, N_l, E_l, O_l, A_l, C_l, Age_l, Com_l]$$

As a result of the regression analysis in Juran (1997), we have a regression equation that calculates the value of each processing time as a linear combination of this vector:

$$X_{jkl} = b_{oj} + \vec{b}_j V_{kl} + \mathbf{e}_{jkl}$$

X_{jkl} = the mean of processing time j for work team k, l .

b_{oj} = the constant (intercept) term in the regression equation for j

\vec{b}_j = the 14-dimensional vector of regression coefficients for j

V_{kl} = the 14-dimensional vector of team demographic and personality attributes.

\mathbf{e}_{jkl} = a random error term.

Using this model, we generated 100 teams of workers, expressed as 14-dimensional vectors. Each element of the vector V_{kl} , was sampled at random, using the demographic prevalence of these characteristics in the subject labor pool. Then processing times were drawn from a different distribution for each pair of workers:

$$D_{jkl} \sim PN(\bar{X}_{jkl}, s_{jkl})$$

Where $\bar{X}_{jkl} = b_{oj} + b_j V_{kl}$, and s_{jkl} is the standard deviation of the estimate from the regression equation for j . We also ran this simulation 100 times, with 100 different realizations of V_{kl} , but using the same random number streams for the error term as before. The results of this experiment appear in

⁵The elements of the vector V are, for each worker, five dimensions of personality (N, E, O, A, and C), plus the worker's age (Age) and degree of computer familiarity (Com). The five personality dimensions are Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C), also known as the "Big Five". These dimensions are measured using the NEO Personality Inventory-Revised (NEO PI-R), an instrument developed by Costa and McCrae (1992). For a general discussion of the Big Five, see Goldberg (1990); for a critique of NEO PI-R, see Keyser and Sweetland (1993) or Leong and Dollinger (1994). The NEO PI-R is a copyrighted publication of Psychological Assessment Resources, Odessa, FL.

Figure 6. This distribution is even wider than the one from Experiment 3 (and much more like the distribution from the lab data), reflecting an even more accurate assessment of the real system’s performance variability.

3. Modeling Individual Workers

3.1 Experiment 5 - Two Specific Workers

In this experiment we make 100 runs of the simulation model, using the same team demographic and personality vector V_{kl} for all 100 runs. The parameters for the processing time distributions are those associated with a particular team instead of those associated with the entire subject pool. This set of parameters chosen represents “team 54” (i.e. the 54th run from Experiment 4), chosen randomly.

	N	E	O	A	C	Age	Com
Worker 1	75	88	107	132	149	404	1
Worker 2	69	64	143	127	127	316	0

Table 1 Characteristics of Workers in Experiment 5

Notice that this distribution is much narrower than those from Experiments 2 and 3 (Figure 7); this reflects the fact that uncertainty regarding the performance of the system is greatly reduced when we know specifically who the workers in the team are. This experiment has the low variability associated with Experiment 1; however, the low variability here is not due to assuming constant known performance parameters across workers but due to the fact that these parameters are for a particular pair of workers. Note that the distribution is not centered at the same place as any of the other performance distributions as expected. We can conclude that if the work team is known in advance, this information lets us get a much more accurate estimate of production cell performance.

3.2 Experiment 6 - Two Generic Workers

There is another way of looking at Experiment 1. This experiment is similar to Experiment 5, except that each element of the 14-dimensional vector of team personality and demographic attributes is set equal to its estimated mean value within the pool of subjects in the lab experiment. Instead of the two

randomly selected workers used in Experiment 5, we have two “vanilla” workers, who are average in every way. As expected, substituting average values for all the demographic and personality attributes into the regression equation for the processing times gives the same result as averaging these parameters before using them in the simulations (as in Experiment 1).

As was the case in Experiment 5, this experiment poorly reflects the randomness in the production cell’s actual performance (see Figure 8). In effect, this experiment is a complicated way to replicate Experiment 1; despite the complex dependency relationships between demographic factors and processing times, when we set all of the independent variables equal to their estimated population means, we end up with our original, naive model.

4. Conclusions and Suggestions for Further Research

The results of the six experiments are summarized in Table 2. Recall that the variable being measured is the number of jobs per hour processed by a team of two workers. As was the case in the histograms (Figures 2 to 8), the data have been converted into orders per hour.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Lab Data
Mean (Orders per Hour)	28.38	28.37	28.77	28.77	55.52	28.37	28.25
Std. Dev	0.15	0.15	5.04	7.15	0.22	0.15	7.20
C.V.	0.53%	0.51%	17.52%	24.84%	0.40%	0.53%	25.48%
Chi Square (5 <i>d.f.</i>)	129.23	129.23	19.80	5.24			
<i>p</i> -value	0.000	0.000	0.001	0.387			
Result of Chi-Square Test	Rejected	Rejected	Rejected	Not Rejected			

Table 2. Summary Statistics from Simulation Experiments

Chi Square statistics and their associated *p*-values are provided for Experiments 1, 2, 3, and 4 to test the hypothesis that these distributions of orders per hour are the same as the “real” distribution observed in the behavioral lab. From the observed significance levels, this hypothesis would most likely be rejected for Experiments 1, 2, and 3 but not rejected for Experiment 4.

Assuming that operations modelers would want to make predictions about the performance of a proposed system on the basis of these simulation experiments, it is clear that the performance we would predict is quite different depending on how we take into account information about the workers. In

Experiments 1 and 2, the length of the run is sufficient to average out almost all of the effects of randomness. While each run has a different random number seed, they all end up producing almost exactly the same number of orders (see the coefficient of variation, or C.V., in Table 2; about one half of one percent).”

In Experiment 3, where each pair of workers has a different set of processing time parameters, we observe a much larger dispersion in performance results. This could be of critical interest to managers in certain situations. Consider, for example, a manager who stands to lose money if the team produces fewer than 1000 orders per 40-hour week. The results of Experiments 1 and 2 suggest that the system will almost always produce more than 1000 orders per week, but the results of Experiment 3 indicate that the system will produce fewer than 1000 orders per week about 23% of the time.

In Experiment 4, demographic data are randomly generated and used to predict processing time distributions on the basis of our regression models. We observe that Experiment 4 approximates most closely the behavior of the system with human workers (compare the coefficients of variation).

These four experiments represent four approaches to modeling a specific system, each with its own trade-off between reality and modeling convenience. The result of including more sources of variation in our input parameters is a model that reflects more accurately the real variability that comes from having different workers operating the system.

Experiments 5 and 6, by contrast, indicate that with two known workers in specific positions the production cell’s performance can be predicted with much greater precision. Just as the models used in Experiments 1 and 2 predict an unrealistically narrow range of performance for a system in which workers are drawn from a diverse pool, the models in Experiments 3 and 4 predict an unrealistically broad range of performance for a system with two specific workers whose demographic and personality characteristics are known in advance.

The differences in these experiments argue against trying to represent differences between workers by using pooled parameter estimates. This technique results in an unrealistic simulated system in

which the variability of processing times within each worker is underestimated. Differences in workers' mean processing rates can cause blocking and starving in tightly-coupled systems because of worker mismatches; underestimating variability will cause the models to underestimate congestion and thus be overly optimistic in predicting system performance.

In general, as more detailed information is introduced into a model, more accurate predictions of how the real system might behave become possible. The techniques demonstrated here indicate several possible levels of detailed information with respect to worker attributes that might be included in a simulation model, each of which has implications for the model's validity. The model in Experiment 4 contains the most detail, and is also the one whose performance most closely represents the performance of the behavioral lab.

This is not to say that this level of detail is always necessary or even useful, but that it is an option to be considered. As always, the prudent modeler must weigh the benefits of a model which closely mimics reality against the costs and complexity involved in designing and analyzing such a model, as well as the costs in collecting the data necessary for the approaches described here. In systems where the predictive validity of worker characteristics is established with respect to the system's performance (as is the case in the two-worker, three-machine behavioral lab), the techniques illustrated here offer a means for improving the fidelity of our models to the systems they represent.

The experiments described here deal only with different approaches to calculating the input parameters for processing time distributions; no attempt is made to model the dynamics by which these parameters might change over time or in response to different system states. However, some of the results in Juran (1997) suggest that such state and time dependent behavior does occur and these changes are also characteristic of individual workers. Examples include worker fatigue or faster processing when a workers input buffer becomes full. The methodology for modeling state-dependent and time-dependent worker behavior needs further development.

Finally, it should be noted that this paper offers only narrow suggestions of what should become a broader research field within the scope of operations management. For example, there is a fertile set of

research questions surrounding the implications of skilled workers, who may be part of a relatively small “pool”. One possible direction for future research is to examine whether a smaller pool of workers means less dampening effect and therefore higher performance variability, or whether it is possible to recruit and select these few workers from a large pool, and thus reduce variability.

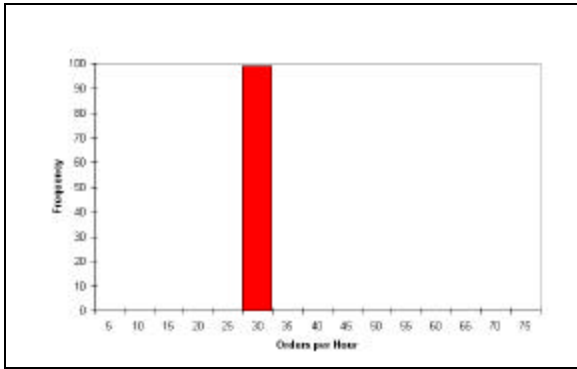


Figure 3. Histogram of Experiment 1 Results

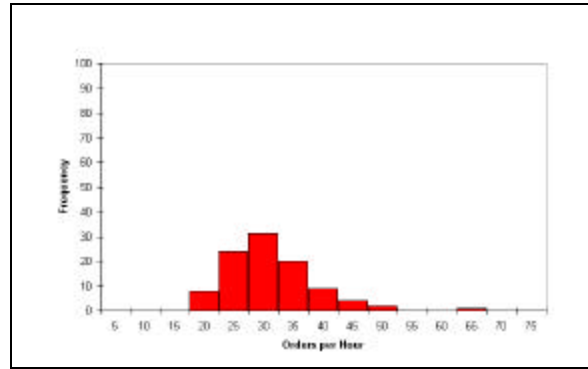


Figure 6. Histogram of Experiment 4 Results

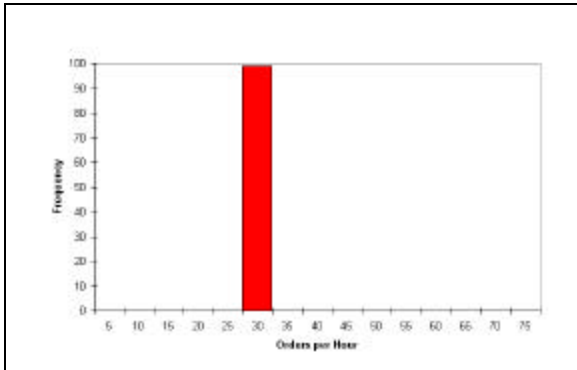


Figure 4. Histogram of Experiment 2 Results

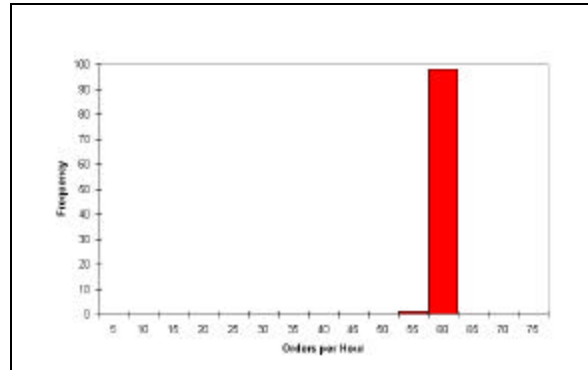


Figure 7. Histogram of Experiment 5 Results

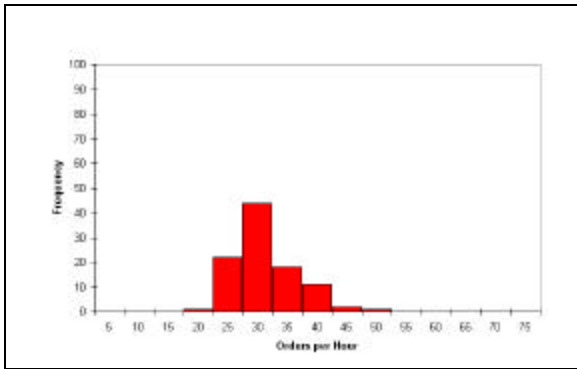


Figure 5. Histogram of Experiment 3 Results

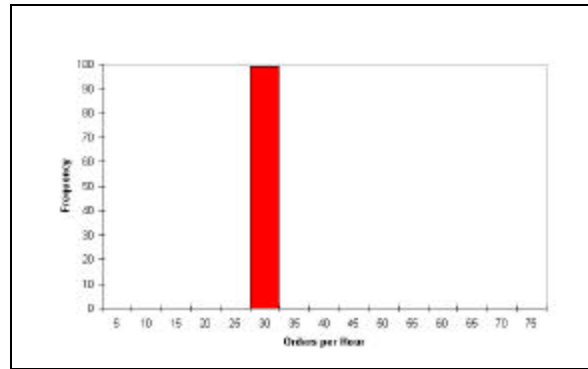


Figure 8. Histogram of Experiment 6 Results

References

- Banks, J. and J.S. Carson II (1984). *Discrete Event Simulation*. Englewood Cliffs, NJ: Prentice-Hall.
- Barrick, Murray R. and Michael K. Mount (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology* 44 (1), 1-26.
- Barry, Bruce, and Greg L. Stewart (1997). Composition, Process, and Performance in Self-Managed Groups: The Role of Personality. *Journal of Applied Psychology* 82 (1) 62-78.
- Bartholdi, John J. III and Donald D. Eisenstein (1996). A Production Line that Balances Itself. *Operations Research* 44 (1), 21-34.
- Bischak, Diane P. (1996). Performance of a Manufacturing Module with Moving Workers. *IIE Transactions* 28, 9.
- Conway, Richard, William Maxwell, John O. McClain, and L. Joseph Thomas (1988). The role of Work-in-process inventory in serial production lines. *Operations Research* 36 2 229-241.
- Costa, Paul T. Jr., and Robert R. McCrae (1992). *Revised NEO Personality Inventory (NEO PI-R)*. Odessa, FL: Psychological Assessment Resources, Inc.
- Doerr, K. H., T. R. Mitchell, T. D. Klastorin, and K. A. Brown (1996). The Impact of Material Flow Policies and Goals on Job Outcomes. *Journal of Applied Psychology* 81, 142-152.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995). *Bayesian Data Analysis*, London: Chapman and Hall
- Goldberg, Lewis R. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology* 59, 1216-1229.
- Gross, Donald and Carl M. Harris (1985). *Fundamentals of queueing theory*. New York: John Wiley and Sons, Inc.
- Helton, J. C. (1996) Computational Structure of a Performance Assessment Involving Stochastic and Subjective Uncertainty. 1996 Winter Simulation Conference.
- Juran, David C. (1997). *The Simulation of Self-Directed Production Systems: Incorporating Human Factors*. Ph.D. Dissertation. Cornell University, Ithaca, NY.
- Juran, David. C., Kenneth L. Schultz, Tove H. Hammer, and John O. McClain (1996). *Personality and the Performance of Production Systems*. Cornell University Johnson Graduate School of Management Working Paper 96-04, Ithaca, NY.

- Kelton, W. David (1984). Winter Simulation Conference Proceedings, 91--95.
- Kelton, W. David, Randall P. Sadowski, and Deborah A. Sadowski (1998). *Simulation with Arena*. Boston, Mass. McGraw-Hill.
- Kempf, Karl G. (1996). Simulating Semiconductor Manufacturing Systems: Successes, Failures, and Deep Questions. Keynote Speech, Proceedings of the 1996 Winter Simulation Conference, Ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain.
- Keyser, Daniel J. and Richard C. Sweetland (1993). *Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business* (2nd ed.) Kansas City: Test Corporation of America. 158.
- Law, Averill M. and W. David Kelton (1991). *Simulation Modeling and Analysis*. New York: McGraw-Hill.
- Lawler, E. E. (1976). New Approaches to Pay: Innovations that Work. *Personnel* 53, 11-23.
- Leong, Frederick T. L., and Stephen J. Dollinger, Jr. (1994). NEO Personality Inventory-Revised, in Daniel J. Keyser and Richard C. Sweetland (Eds.) *Test Critiques, Volume VIII*. Austin, TX: Pro-Ed. 527-39.
- Lin, Neng-Pai, and Hung-Chang Chu (1998). The Effects of Server's Personality on Service Quality. National Taiwan University working paper.
- Mount, Michael K. and Murray R. Barrick (1995). The Big Five Personality Dimensions: Implications for Research and Practice in Human Resources Management. *Research in Personnel and Human Resources Management* 13 153-200.
- Ones, Deniz S., Chockalingam Viswesvaran, and Frank L. Schmidt (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology* 78 (4) 679-703.
- Pegden, C. Dennis, Robert E. Shannon, and Randall P. Sadowski (1995). *Introduction to Simulation Using SIMAN*. New York: McGraw-Hill.
- Schmeiser, Bruce (1999) *Advanced Input Modeling for Simulation Experimentation*. 1999 Winter Simulation Conference.
- Schruben, Lee W. (1992). *Sigma: A Graphical Simulation Modeling Program*. San Francisco: The Scientific Press.
- Schultz, Kenneth L., David C. Juran, John W. Boudreau, John O. McClain, and L. Joseph Thomas (1998). Modeling and Worker Motivation in JIT Production Systems. *Management Science* Vol. 44, No. 12, Part 1 of 2, December 1998 (1595-1607).

Schultz, Kenneth L., David C. Juran, and John W. Boudreau (1999). The Effects of JIT on the Development of Productivity Norms. *Management Science* Vol. 45, No. 12, December 1999 (1664-1678).

Zavadlav, E., John O. McClain, and L. J. Thomas (1996). Self-Buffering, Self-Balancing, Self-Flushing Production Lines. *Management Science* 42 (8), 1151-1164.

The authors gratefully acknowledge the support of the National Science Foundation (Grant DMII-9713549), the Semiconductor Research Corporation (Contract 97-FJ-490), the Clifford H. Whitcomb Faculty Fellowship, and Psychological Assessment Resources, Inc. David Ruppert and four anonymous referees provided helpful comments on an early draft of this paper.