Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Operations Management 22 (2004) 355–367
JOURNAL OF OPERATIONS MANAGEMENT
www.elsevier.com/locate/dsw

Using worker personality and demographic information to improve system performance prediction

David C. Juran [a,b,*], Lee W. Schruben [c]

[a] *Graduate School of Business, Uris Hall, Columbia University, New York, NY 10027, USA*
[b] *Stern School of Business, Tisch Hall, New York University, New York, NY 10012, USA*
[c] *Department of Industrial Engineering and Operations Research, Etcheverry Hall, University of California, Berkeley, CA 94720-1777, USA*

Abstract

This paper presents an approach to modeling workers where human performance has a significant impact on system productivity. Highly technical industries such as semiconductor manufacturing and service industries like banking are relying on fewer but more skilled workers. In these systems, productivity depends on worker availability and organization; therefore, modeling system performance may require accurate representations of individual worker behavior. This paper examines the tradeoffs in including information about the demographics and personalities of workers in system performance simulation models. A series of actual and simulated experiments in which personality and demographic data are used in different ways to model the performance of a team of workers is reported. Significant differences are found in predicted system performance demonstrating that model validity depends 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 tasks.

Keywords: 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 (National Academy of Sciences, 1999). As this trend continues, productivity will become more dependent on the availability and organization of vital human resources. Simulation models of these systems may need to contain accurate representations of individual 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 are often an important consideration. As the politicians from high labor cost regions might want us to believe, prevailing local labor rates may not tell the whole story. The availability of educated, well-trained, motivated and honest workers might be more important than the average labor cost in a given region. However, unless operations models account for these human characteristics, we may not be able to quantify how much they are worth.

* Corresponding author.
E-mail address: dj114@columbia.edu (D.C. Juran).

0272-6963/$ – see front matter
doi:10.1016/j.jom.2004.05.003

Researchers have addressed this modeling issue using various combinations of psychological and operations research methodologies. Some work in organizational science examines the effects of psychological independent variables on psychological variables, such as work group processes (see, for example, Barry and Stewart, 1997). Researchers in operations have examined the effects of structural system variables on system performance measures such as throughput (e.g. Conway et al., 1988).

Interesting discoveries have been made through collaboration between 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 interdisciplinary 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 (increased processing time variability).

This paper focuses on modeling effects of psychological variables on system productivity. The existence of an effect, once established, provides some rationale for including that effect in an operational model. This paper provides a modeling methodology for doing so.

1.1. Parameters and estimators

Most simulation modeling textbooks caution their readers to be wary of replacing a random variable with a constant. Replacing an input random variable in a production system model with an estimate of its expected value can cause the variability in the simulation output to be a gross underestimation, thereby overestimating the productivity of the system being simulated.

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 when running simulation experiments. Kelton (1984) and Banks and Carson (1984), for example, provide very 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)

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. In all these situations, the authors are indirectly recommending replacing a random variable (parameter estimators) with constants (the parameter estimates). Obviously, if additional real world data were collected the values of the parameter estimates would change. This uncertainty in parameter values is almost universally ignored in simulation practice.

Kelton (1984) provides a methodology for conducting sensitivity analysis to different methods of parameter estimation using confidence intervals. While methods for parameter sensitivity analysis are important, they do not resolve the basic problem of replacing random parameter estimators with the constant values, such as the maximum-likelihood estimates of the unknown parameters.

Kelton et al. (1998) state: "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), where he discusses the differences between Kelton's (1996) stochastic and subjective uncertainty, but which also concludes "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".

1.2. Approaches to recognizing randomness

There are several ways to account for the randomness in parameter estimators instead of ignoring it. We use the specific example of worker task time parameters to illustrate the general problem of addressing the variability of input parameter estimators. Later in the paper, we will explore the tradeoffs in including information about the demographics and personalities of workers in system performance simulation models. 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. We hope the lessons learned in this study will help researchers begin to model machines and people with "equal fidelity" as advocated in (Kempf, 1996).

We start by considering the output of a simulation experiment as being composed of two basic elements: (a) the model input, here represented by the vector $\boldsymbol{X}$, and (b) the simulation model itself, here represented by S. On the most basic level, then, the results of running a simulation experiment, Y, can be represented as following:

$$Y = S(\boldsymbol{X}) \tag{1}$$

We further assume that $\boldsymbol{X}$ has some known distribution F, with a set of parameters, P. Using the distribution specified by P, some method of randomization U—typically a function of a pseudo-random number stream—is employed to perform simulation experiments.

$$Y = S[F(P, U)] \tag{2}$$

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

$$\hat{P} = G(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N) \tag{3}$$

We propose several extensions to this basic method for modeling workers, some of which will have significant advantages with respect to the validity of the simulated system.

1.2.1. Randomly selected worker teams

Note that, although all of the distribution parameters in the conventional approach are estimated from sample data (and therefore subject to sampling error), their values are treated as correct and known constants, ignoring any possible difference between P and $\hat{P}$. Instead, we might take a Bayesian statistical approach and move a layer deeper by modeling the input parameters as random variables, which themselves have parameterized prior distributions (Gelman et al., 1995). In systems with human entities, one source of variability is variability across teams of workers. In Section 2.2 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.

1.2.2. Using demographic information

While the previous approach acknowledges variability across samples of worker teams, it ignores demographic information, which may help explain 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 random variables, it is also possible to collect other information about the entities (here workers) being studied. If sufficient reason exists to consider this set of other information (represented by the vector $\boldsymbol{V}$) to be useful in explaining variability in performance[1], we might consider treating the observed performance of certain entities as a dependent variable in a regression model such as:

$$\boldsymbol{X} = \boldsymbol{C}_1\boldsymbol{V}_1 + \boldsymbol{C}_2\boldsymbol{V}_2 + \boldsymbol{C}_3\boldsymbol{V}_3 + \cdots + \boldsymbol{C}_m\boldsymbol{V}_m \tag{4}$$

where $\boldsymbol{V}$ is an m-dimensional vector of attributes that describes the entity, and $\boldsymbol{C}$ is a random vector of the effects on $\boldsymbol{X}$ associated with the attributes $\boldsymbol{V}$, where $\boldsymbol{X}_i$ is the ith observation of the entity's performance. $\boldsymbol{C}$ is estimated using regression analysis, using empirical observations of actual entities. Using empirical

[1] Meta-analytic work among personality psychologists (Barrick and Mount, 1991; Ones and Schmidt, 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).

distributions of the attributes $\boldsymbol{V}$ among the population of real system entities, we can simulate the selection of different entities (teams of workers in our application), and model their performance taking into account the effects of their individual attributes (personality and demographic characteristics in our example). We illustrate this method in Section 2.3.

Some of the elements of $\boldsymbol{V}$ 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 illustrated in Section 2.3). We can then use demographic information to estimate the distributions of the elements of $\boldsymbol{V}$ in the entity population (here the potential labor pool) on. The $\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \ldots, \boldsymbol{C}_m$ 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 estimating when we do this, but we may have the advantage of having much more information from population demographics to estimate M. With this approach, we can infer information about the distribution of $\boldsymbol{V}$ and include knowledge of the randomness in the worker population in our labor pool into our model. This allows us to account for this variability in our sample and in our parameter estimators, and, finally, include this source of randomness in our simulated system performance, Y.

The basic philosophy employed here is two-fold: first, we want to account for randomness in the parameters in the simulation model. Second, we attempt to exploit other sources of empirical knowledge about the entity population (labor pool) from which the data were collected to better estimate this source of randomness.

Several approaches to modeling individual worker differences are demonstrated in the next section using these extensions of standard simulation methodology. Three 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 detail is worth the extra effort or not ultimately depends on the stakes and risks 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 where randomness in this cell may greatly influence the performance of the entire system. Therefore, we wish to model accurately the true 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 are trying to develop a simulation method that will accurately replicate the behavior of the actual system in the lab (operated by human workers).[2] This behavior is not optimal in the sense that it represents the "best" possible configuration of the system. The behavior of the system in the lab is, therefore, a target for us to simulate. A "good" model is one that reflects the actual random behavior of the lab system.

Using the results of the behavioral experiment as our target, we perform three simulation experiments, as described in the following sections. Experiment 1 is intended to be representative of conventional simulation modeling practice; all "workers" are assumed to have the same underlying distributions of processing times. In Experiment 2, each team has a randomly re-sampled set of processing time parameters. In Experiment 3, we vary the processing time parameters of the various teams of workers, explicitly modeling them as dependent variables driven by variation in demographic variables.

Fig. 1 illustrates the cell as simulated and as it was set up in the behavioral laboratory. In this cell,

[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).

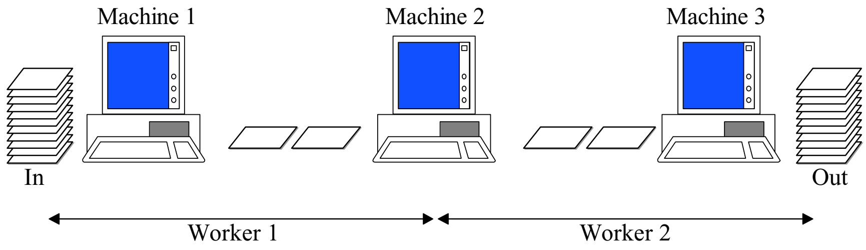

Fig. 1. The laboratory factory.

two-worker teams process jobs on three machines arranged in a serial line. Jobs (or "orders") 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 centers or order fulfillment centers. The actual tasks involved reading, checking, and entering information on a computer keyboard.

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. Fig. 2 is a histogram of the observed cell performance using actual manufacturing workers under carefully measured conditions; details are 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 control for location sampling error so differences between the two-line positions cannot be attributed to having used different pools of workers in the two positions.

1.4. Creating and validating the simulation

A simulation model representing the production cell was wrtten written in C code created using SIGMA (Schruben and Schruben, 2001). In the development of our behavioral laboratory, a concurrently running parallel simulation model of the system a was maintained. The methodology presented here is not dependent on the use of SIGMA, but it did allow full low-level control of the simulation event sequencing that we needed. It was suggested by a reviewer that one might be able to use "human-in-the-loop" software[3], which is an interesting idea worth further consideration.

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

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

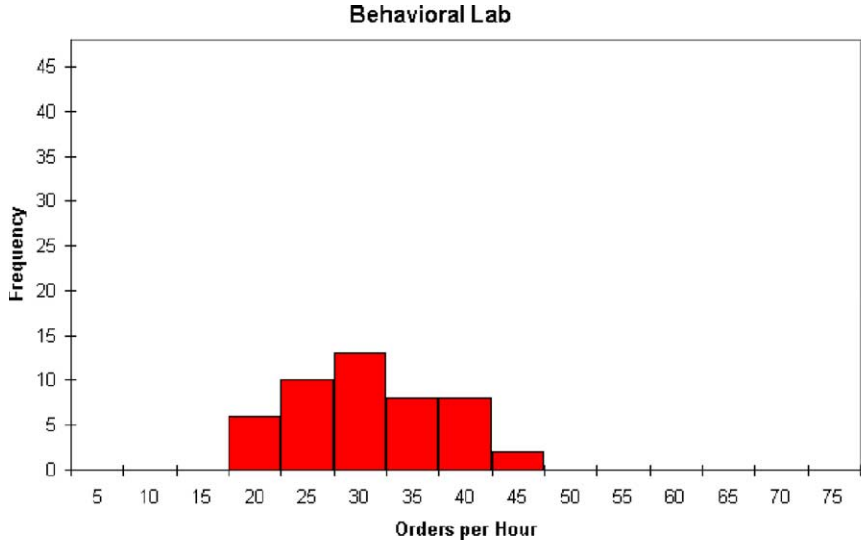

Fig. 2. Histogram of behavioral lab results.

corresponding times in the simulation model. These actual event delay times were used to schedule 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. Then the logic in the model was corrected and the process continued. In this sense, the real system trained the simulation while the two ran together.

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 the event sequences of 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 could be then automated to shadow every laboratory experiment exactly. This was fully monitored throughout the experiments with the real system.

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 since they 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 in about 1% of the generated times. We will denote the conditional truncated-normal[4] probability distributions with mean *m* and standard deviation *s* as *PN*(*m*, *s*)—for “positive normal.” 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 times for 10 input variables: Worker 1 Set-Ups, Worker 1 Processing Items, Worker 1 Take-Downs, Worker 1 Moves, Worker 1 Bumps, Worker 2 Set-Ups, Worker 2 Processing Items, Worker 2 Take-Downs, Worker 2 Moves, and Worker 2 Bumps. There are, therefore, 20 input parameters required for each simulation run (a mean and standard deviation for each of 10 *PN* distributions). The experiments in the remainder of this paper focus on different ways of modeling these 20 parameters. Using the data collected in the behavioral lab, we modeled these parameters at different levels of detail in the simulations.

Every simulation replication in all our experiments was run for 10 million time units or for roughly one workweek. (Real system data traces had to be collected using discrete time units. We chose the time measurement unit of 1/60 s so that 10 million time units = 46.3 h.) The intent of the simulation was to imitate the dynamic behavior of the real system, and therefore, both the simulation runs and laboratory system experiments were started with the system empty. This turned out not to be important at all, as both the real and simulated systems warmed up very rapidly since the queues in both were constrained by the production rules to contain no more than two jobs. Each experiment consisted of 100 simulation runs. The run lengths and numbers of runs in each experiment were somewhat arbitrarily chosen, with the aim of providing sufficiently large samples at a small enough time scale to control these sources of simulation sampling error.

As in the behavioral lab, we kept track of the number of orders processed by the team during each run as the primary measure of performance. All simulation experiments reported in this paper consisted of the same number of independently seeded runs (1 0 0), each of which had the same duration (10 million time units).

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

2. Modeling worker populations

In this section, we will look at several 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. We use the term validity here to mean the degree to which the simulation model approximates the actual performance of the target system. In this specific context, validity means the degree to which the numbers of orders processed by teams in the simulation model approximate the actual observed values from the behavioral lab as measured by characteristics of their probability distribution such as means, variances, and goodness-of-fit results. Perhaps, more remarkable than these statistical comparisons are the visual comparisons of histograms of system and model performance.

2.1. Experiment 1—conventional methodology

In this first experiment, we used a methodology to model our production cell that is intended to be representative of 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 applied here was 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 is 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 as follows:

$$D_{jkl} \sim PN(\hat{\mu}_j, \hat{\sigma}_j) \quad (5)$$

D_{jkl} is the random time for sub-activity j ε (Worker 1 Set-Up, Worker 1 Processing Item, Worker 1 Take-Down, Worker 1 Move, Worker 1 Bump, Worker 2 Set-Up, Worker 2 Processing 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. $\hat{\mu}_j$ is the sample mean time (estimate of μ_j) for j observed in the behavioral laboratory (including all workers in all runs), and $\hat{\sigma}_j$ (estimate of σ_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 each 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. Each run consisted of 10 million simulated time units. A histogram of the results (converted into orders per hour, to be comparable with the behavioral lab results in Fig. 2) appears in Fig. 3. This is a very "tight" distribution; using the same scale as Fig. 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 week-long simulation run is very likely to result in an estimated population mean that is very close to the true population mean.

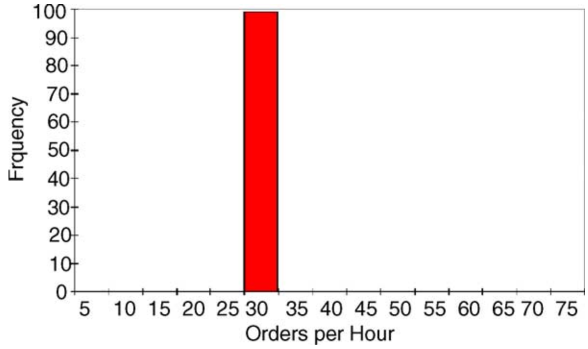

Fig. 3. Histogram of Experiment 1 results.

The main observation here 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 underestimation of a production cell's variability may result in overestimating overall system performance by tending to underestimate the times that individual items spend in queues waiting for a worker.

2.2. Experiment 2—randomly selected teams

This approach is an attempt to account for the fact that we do not know the "true" parameter values for the delay time distributions. 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 the following:

$$D_{jkl} \sim PN(\bar{X}_j, s_j) \quad (6)$$

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 the values of the maximum-likelihood estimators $\bar{X}_j$ and s_j in the simulations 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 ith observation of the inter-event delay j for team kl. First, the sample mean and variance are computed from the data for each team as following:

$$\bar{y}_{jkl} = \frac{1}{n_{jkl}} \sum_{i=1}^{n_{jkl}} y_{ijkl} \quad (7)$$

$$s^2_{jkl} = \frac{1}{n_{jkl}} \sum_{i=1}^{n_{jkl}} (y_{ijkl} - \bar{y}_{jkl})^2 \quad (8)$$

In the above formula s^2_{jkl} 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:

$$\bar{s}^2_j = \frac{1}{n_j} \sum_{kl=1}^{N} s^2_{jkl} \quad (9)$$

$$\tau^2_j = \frac{1}{n_j} \sum_{kl=1}^{N} (s^2_{jkl} - \bar{s}^2_j)^2 \quad (10)$$

For each simulation run, we simulate selecting a new team kl by sampling $\bar{\boldsymbol{X}}_{jkl}$ from the following:

$$PN\left(\bar{y}_j, \frac{s_j}{\sqrt{n_j}}\right) \quad (11)$$

where $\bar{y}_j$ is the average of the calculation in Eq. (7) over all 48 teams and s^2_j is the square root of the result in Eq. (9). We sample a new variance, s^2_{jkl}, from the following:

$$PN(\bar{s}^2_j, \tau_j) \quad (12)$$

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 the orders per hour from these 100 runs appears in Fig. 4. 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 2 requires replicated (or resampled) 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.

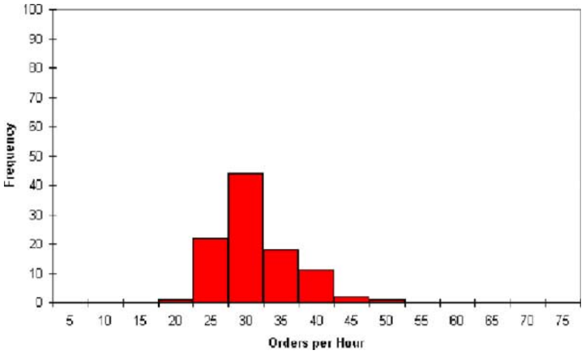

Fig. 4. Histogram of Experiment 2 results.

2.3. Experiment 3—using demographic information

The modeling methodology is next extended 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 explain as much as 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 variables.

$$\boldsymbol{V}_{kl} = [N_k, E_k, O_k, A_k, C_k, \mathrm{Age}_k, \mathrm{Com}_k, N_l, E_l, O_l, A_l, C_l, \mathrm{Age}_l, \mathrm{Com}_l]. \quad (13)$$

The elements of the vector $\boldsymbol{V}$ are, for each worker, the 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.

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:

$$\boldsymbol{X}_{jkl} = \boldsymbol{b}_{0j} + \vec{\boldsymbol{b}}_j \boldsymbol{V}_{kl} + \varepsilon_{jkl} \quad (14)$$

where $\boldsymbol{X}_{jkl}$ is the mean of processing time j for work team k, l; $\boldsymbol{b}_{0j}$ the constant (intercept) term in the regression equation for j; $\vec{\boldsymbol{b}}_j$ the 14-dimensional vector of regression coefficients for j; $\boldsymbol{V}_{kl}$: the 14-dimensional vector of team demographic and personality attributes; and ε_{jkl} the regression random residual error term.

Using this model, we generated 100 teams of workers, expressed as 14-dimensional vectors. Each element of the vector $\boldsymbol{V}_{kl}$ 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:

$$D_{jkl} \sim PN(\bar{\boldsymbol{X}}_{jkl}, s_{jkl}) \quad (15)$$

where $\bar{\boldsymbol{X}}_{jkl}$ is equal to $\boldsymbol{b}_{0j} + \boldsymbol{b}_j \boldsymbol{V}_{kl}$, and s_{jkl} is the standard deviation of the estimate from the regression equation for j. We also ran this simulation 100 times, once each with the 100 different realizations of $\boldsymbol{V}_{kl}$, but using the same random number streams for the error term as before. The results of this experiment appear in Fig. 5. This distribution is more like the distribution from the lab data than the one from Experiment 2 and reflects an even more accurate assessment of the real system's performance variability.

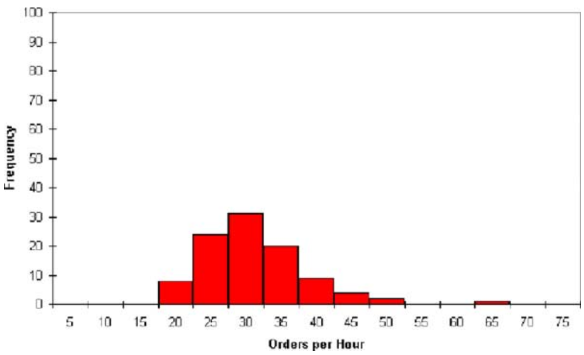

Fig. 5. Histogram of Experiment 3 results.

One can view conventional simulation methodology, our Experiment 1, as a special case of Experiment 3 that includes individual differences across workers, but selects workers who are all identical and "average". In other words, 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 who vary in ways that affect their performance, we have two "vanilla" workers, who are average in every way. As one might expect, when we substituted average values for all the demographic and personality attributes that affect processing times in this manner and re-ran Experiment 3, the results were indistinguishable using average processing time parameters as their estimates in Experiment 1.

3. Conclusions and suggestions for further research

The results of the three experiments are summarized in Table 1. 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 (Figs. 2–5), the data have been converted into orders per hour.

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

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 model the workers. In Experiment 1, following conventional practice of modeling generic workers we get an inaccurate representation of system variability. While each run has a different random number seed, for a week-long simulation they all produce almost exactly the same number of orders as should be expected (see the coefficient

Table 1
Summary statistics from simulation experiments

	Experiment 1	Experiment 2	Experiment 3	Lab data
Mean (orders per hour)	28.38	28.77	28.77	28.25
Sandard deviation	0.15	5.04	7.15	7.20
C.V. (%)	0.53	17.52	24.84	25.48
Chi-square (5 d.f.)	129.23	19.80	5.24	
P-value	0.000	0.001	0.387	
Result of Chi-square test	Rejected	Rejected	Not rejected	

of variation, or C.V., in Table 1 which is about one half of one percent).

In Experiment 2, 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-h week. The results of Experiment 1 suggests that the system will almost always produce more than 1000 orders per week, but the results of Experiment 2 indicates that the system will produce fewer than 1000 orders per week about 23% of the time.

In Experiment 3, demographic data are randomly generated and used to predict processing time distributions on the basis of our regression models. We observe that Experiment 3 closely approximates the behavior of the system with human workers.

These three experiments represent different approaches to modeling a specific system, each with its own trade-off between reality and modeling convenience. The high-level message is obvious: including more sources of variation in our input parameters results in a model that reflects more accurately the real variability that comes from having different workers operating the real system. This is, however, not conventional practice. The surprising impact of modeling individual workers on simulation accuracy found in this study indicates that some of the methods presented here should receive serious consideration.

The differences in these experiments argue against the conventional practice of trying to represent differences between workers by using pooled parameter estimates. That technique results in an unrealistic simulated system in which the variability of processing times across worker teams 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 3 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 issue 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 worker's input buffer becomes full. The methodology for modeling state- and time-dependent worker behavior needs further development.

As an aside, the method we used for creating our simulation model is of some interest in its own right. Here, we had the system itself generate the event logic for the simulation model—the simulation "learned" how to behave from the system by running in parallel with real system time traces to schedule simulated events. This methodology of creating self-simulating systems and the implications on model validity are currently being studied.

Finally, it should be noted that this paper offers only suggestions for simulation modeling of worker performance. The implications about worker behavior in operations management models are much broader.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Foundation (Grant DMII-9713549), the Semiconductor Research Corporation (Contracts FJ-490 and NJ-960), 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.

References

Banks, J., Carson II, J.S., 1984. Discrete Event Simulation. Prentice-Hall, Englewood Cliffs, NJ.

Barrick, Murray R., Mount, Michael K., 1991. The Big Five personality dimensions and job performance: a meta-analysis. Person. Psychol. 44 (1), 1–26.

Barry, Bruce, Stewart, Greg L., 1997. Composition, process, and performance in self-managed groups: the role of personality. J. Appl. Psychol. 82 (1), 62–78.

Bartholdi III, John J., Eisenstein, Donald D., 1996. A production line that balances itself. Operations Res. 44 (1), 21–34.

Bischak, Diane P., 1996. Performance of a manufacturing module with moving workers. IIE Trans. 28, 9.

Conway, Richard, Maxwell, William, McClain, John O., Thomas, L. Joseph., 1988. The role of work-in-process inventory in serial production lines. Operations Res. 36 (2), 229–241.

Costa Paul Jr., T., McCrae Robert, R., 1992. Revised NEO Personality Inventory (NEO PI-R). Psychological Assessment Resources Inc., Odessa, FL.

Doerr, K.H., Mitchell, T.R., Klastorin, T.D., Brown, K.A., 1996. The impact of material flow policies and goals on job outcomes. J. Appl. Psychol. 81, 142–152.

Gelman, Andrew, Carlin, John B., Stern, Hal S., Rubin, Donald B., 1995. Bayesian Data Analysis. Chapman and Hall, London.

Goldberg, Lewis R., 1990. An alternative "Description of Personality": the Big-Five factor structure. J. Personality Social Psychol. 59, 1216–1229.

Juran, David C., 1997. The simulation of self-directed production systems: incorporating human factors. Ph.D. Dissertation. Cornell University, Ithaca, NY.

Kelton, W. David, 1984. Winter Simulation Conference Proceedings. pp. 91–95.

Kelton, W. David, Sadowski, Randall P., Sadowski, Deborah A., 1998. Simulation with Arena. McGraw-Hill, Boston, MA.

Kempf, Karl G., 1996. Simulating semiconductor manufacturing systems: successes, failures, and deep questions. keynote speech. In: Proceedings of the 1996 Winter Simulation Conference, Charnes, J. M., Morrice, D. J., Brunner, D. T., Swain, J. J. (Eds.).

Keyser, Daniel J., Sweetland, Richard C., 1993. Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business, second ed. Test Corporation of America, Kansas City, p. 158.

Law, Averill M., Kelton, W. David, 1991. Simulation Modeling and Analysis. McGraw-Hill, New York.

Lawler, E.E., 1976. New approaches to pay: innovations that work. Personnel 53, 11–23.

Leong, Frederick T.L., Dollinger, Stephen Jr., J., 1994. NEO personality inventory-revised. In: Keyser, Daniel J., Sweetland, Richard C. (Eds.), Test Critiques, vol. VIII. Pro-Ed, Austin, TX, pp. 527–539.

Lin, Neng-Pai, Chu, Hung-Chang, 1998. The Effects of Server's Personality on Service Quality. National Taiwan University working paper.

Mount, Michael K., Barrick, Murray R., 1995. The Big Five personality dimensions: implications for research and practice in human resources management. Res. Person. Hum. Res. Manage. 13, 153–200.

National Academy of Sciences. Commission on Behavioral and Social Sciences and Education. The Changing Nature of Work: Implications for Occupational Analysis. 1999, pp. 263–267.

Ones, Deniz S., Schmidt, Frank L., 1993. Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. J. Appl. Psychol. 78 (4), 679–703.

Schmeiser, Bruce, 1999. Advanced Input Modeling for Simulation Experimentation. Winter Simulation Conference Tutorial, 1999.

Schruben, Donna A., Schruben, L.W., 2001. Graphical Simulation Modeling and Analysis using Sigma, fourth ed. Custom Simulations, Berkeley, CA.

Schultz, Kenneth L., Juran, David C., Boudreau, John W., McClain, John O., Thomas, L. Joseph., 1998. Modeling and worker motivation in JIT production systems. Manage. Sci. 44 (12 (1)), 1595–1607.

Schultz, Kenneth L., Juran, David C., Boudreau, John W., 1999. The effects of JIT on the development of productivity norms. Manage. Sci. 45 (12), 1664–1678.

Zavadlav, E., McClain, John O., Thomas, L.J., 1996. Self-buffering, self-balancing, self-flushing production lines. Manage. Sci. 42 (8), 1151–1164.