Editor's Note: This article and the first two discussions were presented orally at the *Technometrics* session of the 36th Annual Fall Technical Conference held in Philadelphia, Pennsylvania, October 8–9, 1992. The conference was cosponsored by the Chemical and Process Industries, the Statistics Divisions of the American Society for Quality Control, and the Section on Physical and Engineering Sciences of the American Statistical Association.

A Systematic Approach to Planning for a Designed Industrial Experiment

David E. Coleman
Alcoa Laboratories
Alcoa Center, PA 15069

Douglas C. Montgomery
Industrial Engineering Department
Arizona State University
Tempe, AZ 85287

Design of experiments and analysis of data from designed experiments are well-established methodologies in which statisticians are formally trained. Another critical and rarely taught skill is the planning that precedes designing an experiment. This article suggests a set of tools for presenting generic technical issues and experimental features found in industrial experiments. These tools are predesign experiment guide sheets to systematize the planning process and to produce organized written documentation. They also help experimenters discuss complex trade-offs between practical limitations and statistical preferences in the experiment. A case study involving the (computer numerical control) CNC-machining of jet engine impellers is included.

KEY WORDS: Industrial experimental design; Measurement error; Nuisance factors; Statistical consulting.

1. INTRODUCTION

1.1 A Consulting Scenario

Consider the following scenario: An experimenter from the process engineering group comes to you and says: "We are manufacturing impellers that are used in a jet turbine engine. To achieve the claimed performance objectives, we must produce parts with blade profiles that closely match the engineering design requirements. I want to study the effect of different tool vendors and machine set-up parameters on the dimensional variability of the parts produced on the machines in our CNC-machine center."

Many experimental design applications in industry begin with such a statement. It is well recognized that the planning activities that precede the actual experiment are critical to successful solution of the experimenters' problem (e.g., see Box, Hunter, and Hunter 1978; Hahn 1977, 1984; Montgomery 1991; Natrella 1979). Montgomery (1991) presented a seven-step approach for planning experiments, summarized in Table 1. The first three of these steps constitutes the preexperiment planning phase. The detailed, specific activities in this phase are the focus of this article. The emphasis is planning for a screening experiment, or a step in sequential experimentation on an existing product/process, off-line or on-line. Many of the issues addressed, however, also apply to new products/processes or research and development (R&D) and to various additional experimental goals, such as optimization and robustness studies.

1.2 A Gap

It is often said that no experiment goes exactly as planned, and this is true of most industrial experiments. Why? One reason is that statisticians who design experiments with scientists and engineers (the "experimenters") usually have to bridge a gap in knowledge and experience. The consequences of not bridging this gap can be serious.

The statistician's lack of domain knowledge can lead to:

1. Unwarranted assumptions of process stability during experimentation
2. Undesirable combinations of control-variable levels in the design
3. Violation or lack of exploitation of known physical laws
4. Unreasonably large or small designs

Table 1. Steps of Experimentation

1.	Recognition of and statement of the problem
2.*	Choice of factors and levels
3.*	Selection of the response variable(s)
4.	Choice of experimental design
5.	Conduction of the experiment
6.	Data analysis
7.	Conclusions and recommendations

*In some situations, steps 2 and 3 can be reversed.

5. Inappropriate confounding
6. Inadequate measurement precision of responses or factors
7. Unacceptable prediction error
8. Undesirable run order

The experimenter's lack of statistical knowledge can lead to:

1. Inappropriate control-variable settings (e.g., range too small to observe an effect or range so large that irrelevant mechanisms drive the response variable)
2. Misunderstanding of the nature of interaction effects, resulting in unwisely confounded designs
3. Experimental design or results corrupted by measurement error or setting error
4. Inadequate identification of factors to be "held constant" or treated as nuisance factors, causing distorted results
5. Misinterpretation of past experiment results, affecting selection of response variables or control variables and their ranges
6. Lack of appreciation of different levels of experimental error, leading to incorrect tests of significance

This article attempts to help bridge the gap by providing a systematic framework for predesign information gathering and planning. Specifically, we present *guide sheets* to direct this effort. The use of the guide sheets is illustrated through the (computer numerical control) CNC-machinery example briefly presented previously. This article is a consolidation and extension of the discussion by Hahn (1984), Box et al. (1978), Montgomery (1991), Natrella (1979), Bishop, Petersen, and Traysen (1982), and Hoadley and Kettenring (1990).

The guide sheets are designed to be discussed and filled out by a multidisciplinary *experimentation team* consisting of engineers, scientists, technicians/operators, managers, and process experts. These sheets are particularly appropriate for complex experiments and for people with limited experience in designing experiments.

The sheets are intended to encourage the discussion and resolution of generic technical issues needed *before* the experimental design is developed. Hahn (1984) listed most of these issues and made the recommendation, "The major mode of communication between the experimenter and the statistician should be face-to-face discussion. The experimenter should, however, also be encouraged to document as much of the above information as possible ahead of time" (p. 25). Unfortunately, as Hahn observed, "Not all experimenters are willing to prepare initial documentation" (p. 26). Moreover, not all of the relevant issues may be thoroughly thought out—hence, the need for face-to-face discussions, during which, as Hahn advised, "The statistician's major functions are to help structure the problem, to identify important issues and practical constraints, and to indicate the effect of various compromises on the inferences that can be validly drawn for the experimental data" (p. 21).

The guide sheets proposed in this article outline a systematic "script" for the verbal interaction among the people on the experimentation team. When the guide sheets are completed, the team should be well equipped to proceed with the task of designing the experiment, taking into account the needs and constraints thus identified.

2. PREDESIGN MASTER GUIDE SHEET AND SUPPLEMENTARY SHEETS

The guide sheets consist of a "Master Guide Sheet," plus supplementary sheets and two tutorials. These are schematically illustrated in Figure 1. The supplementary sheets are often more convenient for items 3–7.

The Master Guide Sheet is shown in Figure 2. It is stripped of the blank space usually provided to fill in the information. Blank copies will be provided by the authors on request.

Discussion of issues related to different pieces of the Master Guide Sheet and the supplementary sheets follows.

Writing the objective (item 2, Fig. 2) is harder than it appears to most experimenters. Objectives should be (a) unbiased, (b) specific, (c) measurable, and (d) of practical consequence. To be *unbiased*, the experimentation team must encourage participation by knowledgeable and interested people with diverse perspectives. The data will be allowed to speak for themselves. To be *specific and measurable*, the objectives should be detailed and stated so that it is clear whether they have been met. To be of *practical consequence*, there should be something that will be done differently as a result of the outcome of the experiment. This might be a change in R&D direction, a change in process, or a new experiment. Conducting an experiment constitutes an expenditure of resources *for some purpose*.

Thus experimental objectives should not be stated

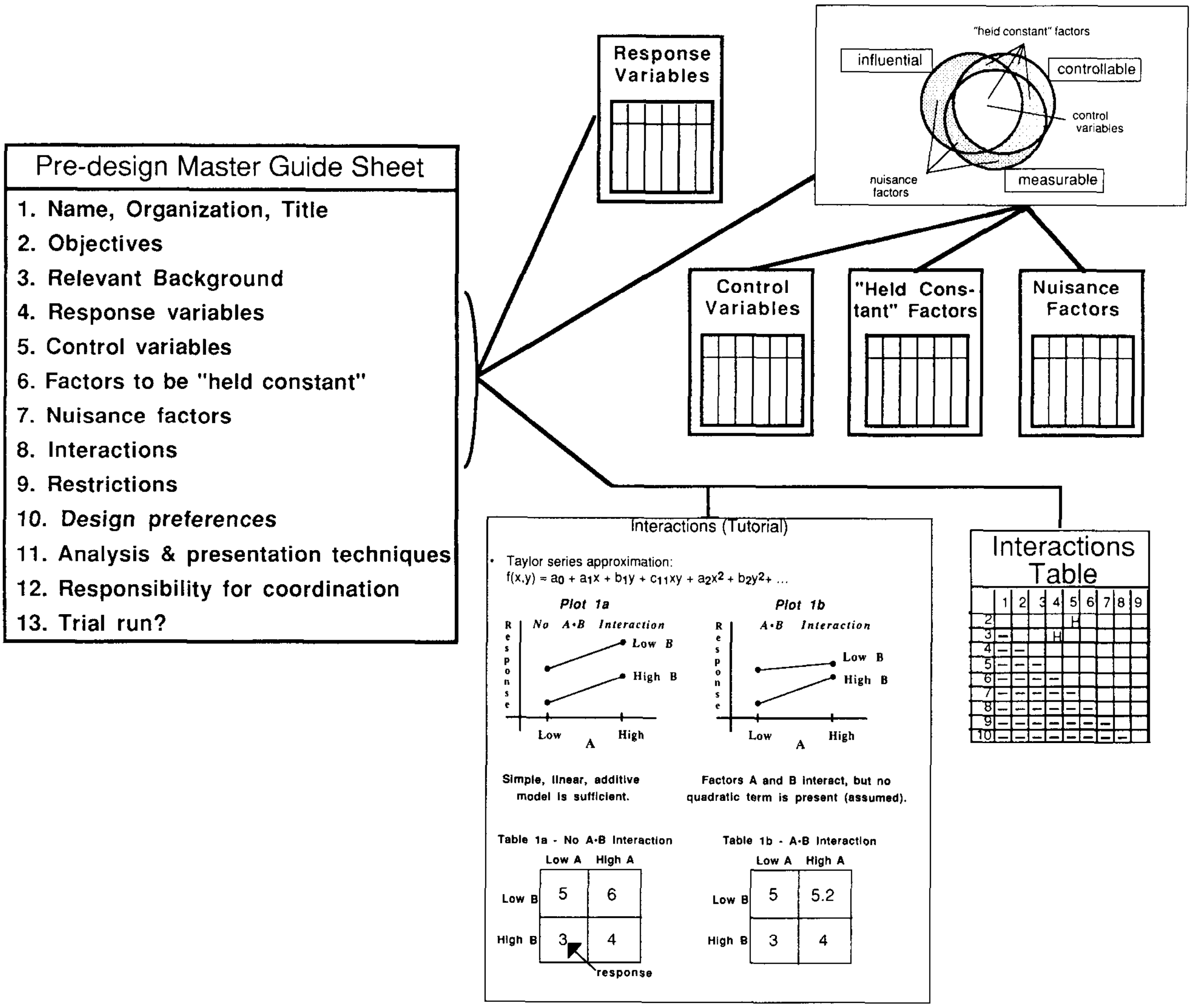

Figure 1. Structure of Predesign Experiment Guide Sheets.

as, "To show that catalyst z14 works better than catalyst d12, if the technician adjusts the electrode voltage just right." A better objective would be: "To quantify the efficiency difference, Δ, between catalysts z14 and d12 for electrode voltages 7, 8, and 9 in the ABC conversion process—and assess statistical significance (compare to 95%) and practical significance ($\Delta > 3\%$), perhaps economically justifying one catalyst over the other."

As Box et al. (1978, p. 15) put it (paraphrased), the statistician or other members of the experimentation team should "ensure that all interested parties agree on the objectives, agree on what criteria will determine that the objectives have been reached, and arrange that, if the objectives change, all interested parties will be made aware of that fact and will agree on the new objectives and criteria." Even experimenters in the physical sciences—who have been trained in the scientific method—sometimes need prodding in this.

The objective of the experiment can be met if the predesign planning is thorough, an appropriate design is selected, the experiment is successfully conducted, the data are analyzed correctly, and the results are effectively reported. By using a systematic approach to predesign planning, there is greater likelihood that the first three conditions will occur. This increases the likelihood of the fourth. Then the experiment is likely to produce its primary product—new knowledge.

2.1 Relevant Background

The relevant background supporting the objectives should include information from previous experiments, routinely collected observational data, physical laws, and expert opinion. The purposes of providing such information are (a) to establish a context for the experiment to clearly understand what new knowledge can be gained; (b) to motivate discussion about the relevant domain knowledge, since

1. Experimenter's Name and Organization: **Brief Title of Experiment:**
2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):
3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?:
4. List: (a) each **response variable**, (b) the normal response variable level at which the process runs, the distribution or range of normal operation, (c) the precision or range to which it can be measured (and how):
5. List: (a) each **control variable**, (b) the normal control variable level at which the process is run, and the distribution or range of normal operation, (c) the precision (s) or range to which it can be set (for the experiment, not ordinary plant operations) and the precision to which it can be measured, (d) the proposed control variable settings, and (e) the predicted effect (at least qualitative) that the settings will have on each response variable:
6. List: (a) each factor to be **"held constant"** in the experiment, (b) its desired level and allowable s or range of variation, (c) the precision or range to which it can measured (and how), (d) how it can be controlled, and (e) its expected impact, if any, on each of the responses:
7. List: (a) each **nuisance factor** (perhaps time-varying), (b) measurement precision, (c) strategy (e.g., blocking, randomization, or selection), and (d) anticipated effect:
8. List and label known or suspected **interactions:**
9. List **restrictions** on the experiment, e.g., ease of changing control variables, methods of data acquisition, materials, duration, number of runs, type of experimental unit (need for a split-plot design), "illegal" or irrelevant experimental regions, limits to randomization, run order, cost of changing a control variable setting, etc.:
10. Give current **design preferences**, if any, and reasons for preference, including blocking and randomization:
11. If possible, propose **analysis and presentation techniques**, e.g., plots, ANOVA, regression, plots, t tests, etc.:
12. Who will be responsible for the coordination of the experiment?
13. Should trial runs be conducted? Why / why not?

Figure 2. Predesign Master Guide Sheet. This guide can be used to help plan and design an experiment. It serves as a checklist to accelerate experimentation and ensures that results are not corrupted for lack of careful planning. Note that it may not be possible to answer all questions completely. If convenient, use the supplementary sheets for 4–8.

such discussion may change the consensus of the group, hence the experiment; and (c) to uncover possible experimental regions of particular interest and others that should be avoided. With this background, we reduce the risks of naive empiricism and duplication of effort.

For the CNC-matching problem introduced earlier, we have the guide sheet shown in Figure 3.

3. RESPONSE VARIABLES

As mentioned previously, items 4–8 on the guide sheet are most conveniently handled using the supplementary sheets. The first one is for response variables, as shown in Table 2.

Response variables come to mind easily for most experimenters, at least superficially; they know what outcomes they want to change—a strength, a failure rate, a concentration, or a yield. What makes a good response variable? The answer to this question is complex. A complete answer is beyond the scope of this article, but some guidelines can be given. A response variable

1. Is preferably a continuous variable. Typically, this will be a variable that reflects the continuum of a physical property, such as weight, temperature, voltage, length, or concentration. Binary and ordinal variables have much less information content—much as the raw values are more informative than histograms that have wide bins. Note that being continuous *with respect to a control variable* may be important. If a response variable has, perhaps, a steep sigmoidal response to a control variable, it is effectively binary as that variable changes. For example,

1. Experimenter's Name and Organization: John Smith, Process Eng. Group
Brief Title of Experiment: CNC Machining Study

2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):

For machined titanium forgings, quantify the effects of tool vendor; shifts in a-axis, x-axis, y-axis, and z-axis; spindle speed; fixture height; feed rate; and spindle position on the average and variability in blade profile for class X impellers, such as shown in Figure 4.

3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?:

(a) Because of tool geometry, x-axis shifts would be expected to produce thinner blades, an undesirable characteristic of the airfoil.

(b) This family of parts has been produced for over 10 years; historical experience indicates that externally reground tools do not perform as well as those from the "internal" vendor (our own regrind operation).

(c) Smith (1987) observed in an internal process engineering study that current spindle speeds and feed rates work well in producing parts that are at the nominal profile required by the engineering drawings - but no study was done of the sensitivity to variations in set-up parameters.

Results of this experiment will be used to determine machine set-up parameters for impeller machining. A robust process is desirable; that is, on-target and low variability performance regardless of which tool vendor is used.

Figure 3. Beginning of Guide Sheet for CNC-Machining Study.

weight of precipitate as a function of catalyst may be near zero for the selected low levels of catalyst and near maximum for the high levels.

2. Should capture, as much as possible, a quantity or quality of interest for the experimental unit. For example, if the experimental unit is an ingot and a response is T = temperature, it may matter whether T is taken at a single point or averaged over a surface region, the entire surface area, or the entire ingot volume.

3. Should be in appropriate units. The units may be absolute, such as pounds, degrees centigrade, or meters. They may be relative units, such as percent of concentration by weight or by volume or proportional deviation from a standard. What is "appropriate" may be determined by an empirical or first-principles model, such as using absolute units in $E = mc^2$, or it may be determined by practical limitations, such as using percent of concentration by weight because the experimental samples are not all the same weight.

4. Should be associated with a target or desirable

Table 2. Response Variables

Response variable (units)	*Normal operating level and range*	*Measurement precision, accuracy—how known?*	*Relationship of response variable to objective*
Blade profile (inches)	Nominal (target) $\pm 1 \times 10^{-3}$ inches to $\pm 2 \times 10^{-3}$ inches at all points	$\sigma_E \approx 1 \times 10^{-5}$ inches from a coordinate measurement machine capability study	Estimate mean absolute difference from target and standard deviation of difference
Surface finish	Smooth to rough (requiring hand finish)	Visual criterion (compare to standards)	Should be as smooth as possible
Surface defect count	Typically 0 to 10	Visual criterion (compare to standards)	Must not be excessive in number or magnitude

condition (which motivates the experiment). Such a comparison might be used to derive "performance measures" from response-variable outcomes. For example, with CNC-machining, blade profile is a response variable, and it is compared to the target profile by computing differences at certain locations. Mean absolute difference and the standard deviation of the differences are performance measures for the various experimental conditions. They can be analyzed separately or by using the standard deviations to compute weights for the mean analysis.

5. Is preferably obtained by nondestructive and nondamaging methods so that repeated measures can be made and measurement error can be quantified.

6. Should not be near a natural boundary. Otherwise, the variable will not discriminate well. For example, it is hard to distinguish a yield of 99.5% from 99.8%, and it is hard to detect and distinguish contamination levels near 0.

7. Preferably has constant variance over the range of experimentation.

There are other important characteristics of response variables that the experimenters may not have considered or communicated to the whole experimentation team. This sheet helps to draw them out: (a) current use, if any (col. 2); (b) ability to measure (col. 3); and (c) the knowledge sought through experimentation (col. 4).

It is helpful to know the current state of use, and if it is unknown, the experimenters are advised to include some trial runs prior to the experiment or "checkpoint" runs during the experiment (perhaps these data have not been previously acquired). The current distribution serves as one of several possible reference distributions for judging the *practical* magnitude of the effects observed. Given a typical standard deviation for a response variable of σ, a low-to-high control-variable effect of $\sigma/2$ may be of no practical significance, but one of 4σ may be important. Another advantage to knowing the current state of use is a check on credibility. Process or design limitations may constrain a response variable to be bounded on one or two sides. An experimental result outside that range may be erroneous or due to an abnormal mechanism (which may, however, be of interest).

Measurement precision (and, in some cases, bias) and how to obtain it (i.e., choice of measurement system or repeated measurements) is a thorn in the flesh for many experimenters. The admonition of Eisenhart (1962) serves as a relevant (if overstated) warning, "until a measurement operation . . . has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all" (p. 162). It has been our experience that many experimenters do not know the state of control nor the precision and bias of most measurement systems measuring a response or a control variable. The measurement systems were not *useless*, they were just of unknown *utility*. Important and poorly understood systems should be evaluated with a measurement capability study, a designed experiment of the measurement process. As a compromise, one might be forced to resort to historical data and experience or weaken the experimental objective to obtain ranking, selection, or a binary response instead of quantification.

The relationship of a response variable to the objective may be direct. An objective may be defined in terms of a response variable—for example, "to quantify the effect that thermal cycle B has on tensile strength measured on customer qualifying tester X." In the case of CNC-machining, a response variable is blade profile (see Fig. 4). This is related to the objective through two measurement-performance indicators—mean absolute difference of blade profile and the target, and standard deviation of the difference. Sometimes a response variable may be a *surrogate* for the true response of interest. This is often the case in destructive testing, in which a standard stress-to-fracture test, for example, represents performance under conditions of use. Another example is yield rate or failure rate, which are inferior responses that often represent where a specification falls relative to a distribution of continuous-scale values (the collection of which provides superior information).

As discussed previously, the relationship of a response variable to the objective may be through *performance measures* that involve a comparison of the response to a target or desirable outcome.

4. CONTROL VARIABLES

As with response variables, most investigators can easily generate a list of candidate control variables.

Control variables can be attribute or continuous. They can be narrowly defined, such as "percent of copper, by weight," or broadly defined, such as "comparably equipped pc: Apple or IBM." In either case, control variables should be explicitly defined.

When discussing potential control variables with experimenters, it may be helpful to anticipate that *held-constant factors* and *nuisance factors* must also be identified. Figure 5 is a Venn diagram that can be used to help select and prioritize candidate factors. It illustrates different categories of factors that affect response variables, based on three key characteristics—magnitude of influence on response variables, degree of controllability, and measurability (e.g., precision and accuracy). Each type of factor is discussed in detail in following sections. A descrip-

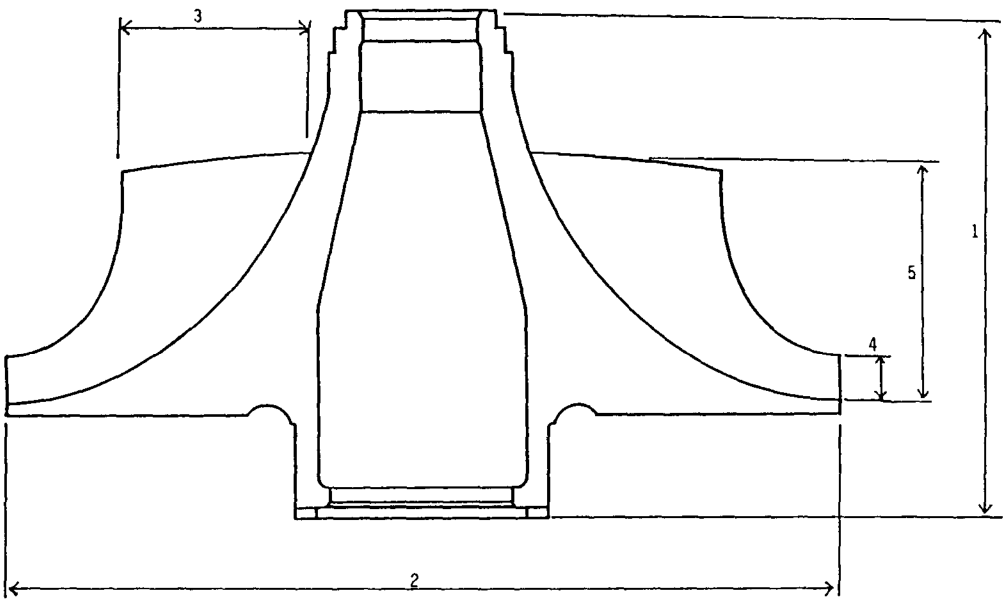

Figure 4. Jet Engine Impeller (side view; z axis is vertical, x axis is horizontal, and y axis is into the page): 1. Height of Wheel; 2. Diameter of Wheel; 3. Inducer Blade Height; 4. Exducer Blade Height; 5. Z Height of Blade.

tion of the diagram is as follows:

1. Control variables are measurable, controllable, and thought to be (very) influential.
2. Held-constant factors are controlled.
3. Nuisance factors are uncontrolled factors (either they cannot be controlled, or they are allowed to vary).

In discussing different variables and factors the team may choose to reassign variables from one group to another, and this is part of the ordinary process for planning a designed experiment. For the CNC-machining problem, the control-variable information was developed as shown in Table 3; those below the space are considered to be of secondary importance.

Similar to the response variables sheet, the control variables sheet solicits information about (a) current use (col. 2), (b) ability to measure and set (col. 3), and (c) knowledge sought through experimentation (cols. 4–5).

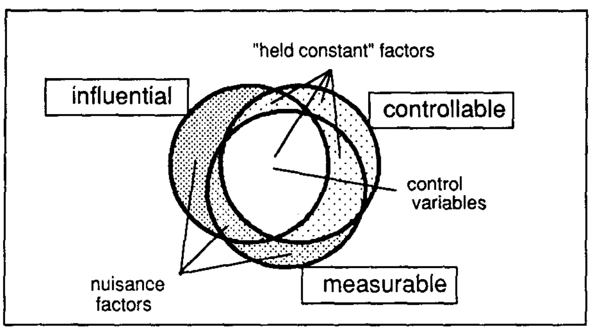

Figure 5. Different Categories of Factors Affecting Response Variables.

4.1 Current Use (col. 2)

There are two reasons it helps to know the allowed ranges and nominal values of control variables under current use. First, the degree to which historical process data can be used to gain relevant knowledge may be revealed. This is discussed in Section 4.2. Second, the experimenter should select a range large enough to produce an observable effect and to span a good proportion of the operating range, yet not choose so great a range that no empirical model can be postulated for the region, as discussed in Section 4.3. In some, less mature experimental situations, there may be no well-defined "current use," in which case trial runs before or during experimentation are helpful—as they are with response variables.

4.2 Ability to Measure and Set (col. 3)

With control variables, there is an additional consideration rarely mentioned in the literature. The experimentation team not only needs to know how *measurements will be obtained* and the precision of measurement, σ_m, but also *how the control variable settings will be obtained* and "setting error," ε_s. These different types of deviation from the ideal have different effects on experimentation. Large σ_m will mean that either errors-in-variables methods will have to be used (e.g., methods that will allow estimation of bias in effects estimates) or, alternatively, many samples will have to be collected for measurement during experimentation to get an acceptably small $\sigma_m/\sqrt{n}$, especially if $|\varepsilon_s|$ is also large. If $|\varepsilon_s|$ is large, traditional, class-variable-based analysis of variance will have to

Table 3. Control Variables

Control variable (units)	*Normal level and range*	*Measurement precision and setting error—how known?*	*Proposed settings, based on predicted effects*	*Predicted effects (for various responses)*
x-axis shift* (inches)	0—.020 inches	.001 inches (experience)	0, .015 inches	Difference ↗
y-axis shift* (inches)	0—.020 inches	.001 inches (experience)	0, .015 inches	Difference ↗
z-axis shift* (inches)	0—.020 inches	.001 inches (experience)	?	Difference ↗
Tool vendor	Internal, external	—	Internal, external	External is more variable
a-axis shift* (degrees)	0—.030 degrees	.001 degrees (guess)	0, .030 degrees	Unknown
Spindle speed (% of nominal)	85—115%	~1% (indicator on control panel)	90%, 110%	None?
Fixture height	0—.025 inches	.002 inches (guess)	0, .015 inches	Unknown
Feed rate (% of nominal)	90—110%	~1% (indicator on control panel)	90%, 110%	None?

*The x, y, and z axes are used to refer to the part *and* the CNC machine. The a axis refers only to the machine.

be replaced by regression analysis. The result of large setting variation may be unwanted aliasing, greater prediction error, violation of experiment constraints, and difficulty conducting split-plot analyses.

Often, one finds that $\sigma_m \approx |\varepsilon_s|$, such as when the measurement system is part of a controller, and equilibrium conditions can be achieved. Measurement precision and setting error are not always comparable, however. For example, $\sigma_m < |\varepsilon_s|$ is not unusual for a continuous-batch mixing process. Suppose that the concentration of constituent A is at 10% and is continuously reduced towards a target of 5%. Batches might be produced with concentrations of 10%, 7%, and 4%. In this case, perhaps $|\varepsilon_s| \approx 1\%$, but a spectrograph may measure with $\sigma_m \approx .1\%$. Another example is a thermostat, which often provides $\sigma_m < |\varepsilon_s|$, especially if it has a "dead zone" in its logic.

Alternatively, one may find $\sigma_m > |\varepsilon_s|$. For example, physical laws may make it possible to accurately *set* gas pressure in a sealed cavity by setting gas temperature, but there may be no precise way to directly *measure* pressure.

4.3 Knowledge Sought Through Experimentation

In the design of experiments classes he teaches at Alcoa, J. S. Hunter gives a rule of thumb for experiments on existing processes. For each control variable, low/high settings should be selected to cause a predicted effect (main effect) for the "key" response variable equal to one standard deviation of its variation in ordinary use, σ_p (if there is "ordinary use"). This is a large enough change in response to have practical consequence and also large enough to likely be detected if measurement error is negligible and the experiment has enough runs. If the rule of thumb is followed, every control variable has "equal opportunity" to affect the response variable.

Naturally, it is harder to suggest such a rule for immature processes. Moreover, other issues and constraints must be taken into account when settings are selected—safety, discreteness of settings, process constraints, ease of changing a setting, and so forth. These are solicited in item 8 of the guide sheet.

Predicted effects for the response variables may be available from the knowledge sources previously listed—theory, experts, and experiments. Quantitative predicted effects are preferable, but experimenters may not be able to provide more than qualitative indications. Even if uncertain, the exercise of attempting to predict the outcome of the experiment before it is run can foster good interaction within the experimentation team and often leads to revised choices of settings. An additional advantage is that the predictions will always be wrong, so it is easier to see what knowledge has been gained through experimentation.

5. HELD-CONSTANT FACTORS

Held-constant factors are controllable factors whose effects are not of interest in this experiment. Most experimenters think in these terms: "For this experiment, I want to study the effect of factors A, B, and C on responses y_1 and y_2, but all other control variables should be held at their nominal settings, and I do not want extraneous factors distorting the results." This sheet was developed to ensure the "all other control variables held at their nominal settings" condition. (The next sheet is used to help ensure that "there are no extraneous factors distorting the results.") For the CNC-machining example, the held-constant factors are as shown in Table 4.

The sheet in Table 4 can force helpful discussion about which factors are adequately controlled and which factors are not. In so doing, it is often necessary to consult experts to help prioritize factors, recommend preexperiment studies to assess control, or develop control strategy.

For example, in the CNC-machining case, this sheet resulted in the recognition of the fact that the machine had to be fully warmed up before cutting any blade forgings. The actual procedure used was to mount the forged blanks on the machine spindles and run a 30-minute cycle without the cutting tool engaged. This would allow all machine parts and the lubricant to reach normal, steady-state operating temperature. The use of a "typical" (i.e., "mid-level") operator and the blocking of the blank forgings by lot number were decisions made for experimental insurance, although neither variable was expected to have important effects. Not that it is not practical or desirable to hold some factors constant. For example, although it might be ideal to have experimental material from only one titanium forging, there may not be enough material within one forging, and forging may interact with experimental variables. The operator's role in this highly automated process is small, and material properties of the blank titanium forgings are carefully controlled because of the criticality of the part.

6. NUISANCE FACTORS

Processes vary over time. Experimental conditions vary over time. "Identical samples" differ. Some variations are innocuous, some are pernicious. Examples include contamination of process fluids over time, equipment wear, build-up of oxides on tools, and so forth. Some of these can be measured and monitored to at least ensure that they are within limits; others must be assessed subjectively by experts; still others are unmeasured. Nuisance factors are not controlled, and are not of primary interest in this experiment. They differ from held-constant factors in that they cannot be deliberately set to a constant level for all experimental units. If the level can be selected for any experimental unit, however, blocking or randomization might be appropriate. If levels cannot be selected (i.e., the levels of the factor are unpredictable, perhaps continuous), then the nuisance factor becomes a covariate in the analysis. If a nuisance factor is not measurable and thought to be very influential, it may also be called an *experimental risk* factor. Such factors can inflate experimental error, making it more difficult to assess the significance of control variables. They can also bias the results. For the CNC-machining example, the nuisance factors are as shown in Table 5.

Experiment designers have a set of passive strategies (randomization, blocking, analysis of covariance, stratified analysis) to reduce the impact of nuisance factors. These strategies can have a major effect on the experimental design. They may be constrained

Table 4. Held-Constant Factors

Factor (units)	*Desired experimental level and allowable range*	*Measurement precision—how known?*	*How to control (in experiment)*	*Anticipated effects*
Type of cutting fluid	Standard type	Not sure, but thought to be adequate	Use one type	None
Temperature of cutting fluid (degrees F.)	100–110°F. when machine is warmed up	1–2° F. (estimate)	Do runs after machine has reached 100°	None
Operator	Several operators normally work in the process	—	Use one "mid-level" operator	None
Titanium forgings	Material properties may vary from unit to unit	Precision of lab tests unknown	Use one lot (or block on forging lot, only if necessary)	Slight

Table 5. Nuisance Factors

Nuisance factor (units)	*Measurement precision—how known?*	*Strategy (e.g., randomization, blocking, etc.)*	*Anticipated effects*
Viscosity of cutting fluid	Standard viscosity	Measure viscosity at start and end	None to slight
Ambient temperature (°F.)	1–2° F. by room thermometer (estimate)	Make runs below 80°F.	Slight, unless very hot weather
Spindle	—	Block or randomize on machine spindle	Spindle-to-spindle variation could be large
Vibration of machine during operation	?	Do not move heavy objects in CNC machine shop	Severe vibration can introduce variation within an impeller

by limits on the number of observations, costs of changing control-variable settings, and logistic considerations. In the CNC-machining example, the only nuisance factor to have potentially serious effects and for which blocking seems appropriate is the machine spindle effect (though it may be necessary to also block on titanium forgings). The machine has four spindles, requiring a design with four blocks or randomizing on all four. Blocking will introduce a bias in the estimates confounded with the blocking variable(s), whereas randomization will inflate the experimental error. The other two factors are dealt with by ensuring that they stay below levels at which problems may be encountered.

7. INTERACTIONS

The interactions sheet is self-explanatory. Unfortunately, the concept of interactions is *not* self-explanatory—even among intelligent, mathematically inclined people in the sciences. Hence, as part of the package of guide sheets, it is helpful to include a tutorial. The graphical portion of the tutorial is presented in Figure 6. An additional, expository description of interactions is sometimes included, but it is not shown here. The interactions table explicitly recognizes only pairwise interactions of linear terms. It provides an opportunity for the experimenters to capture knowledge or speculation that certain pairwise interactions may be present and others are unlikely to be present. This input is helpful when the experiment is later designed—to choose resolution, or more generally to choose which effects should or should not be confounded. Higher order effects may also be important but are not captured in the guide sheets. The interaction sheet for the CNC-machining example is shown in Table 6.

A helpful way to use this matrix is to *avoid* discussing every possible pairwise interaction one at a time but instead use the process of elimination or inclusion; that is, if interactions are generally important, a question can be posed: "Are there any interactions that are arguably *not* present?" If main effects dominate interactions, a question can be posed: "Are there any interactions that must be estimated clear of main effects?" Alternatively, a secret-ballot vote on potentially important interactions can be held among experimenters and other knowledgeable investigators, with each receiving, say, 100 votes to spread among the interactions.

The remaining items are found on the Master Guide Sheet.

8. RESTRICTIONS, PREFERENCES FOR THE DESIGN, ANALYSIS, AND PRESENTATION

Box, Hunter, and others have repeatedly exhorted, "Attention to detail can determine the success or failure of the experiment." Item 8 in Figure 2 is part of heeding that advice. *Theoretical* optimal experimental design and *practical* experimental design are often worlds apart, and restrictions often make the difference. Since a single unknown restriction can render worthless an otherwise well-considered, laboriously developed design, the statistician should encourage experimenters to be quick to put these limitations and pitfalls on the table. In particular, there appears to be a lack of awareness in the applied statistics community of the prevalence of experiments with unidentified split-plot structure. Because it is unidentified (different experimental units used for different parts of the experiment), the analysis is often done incorrectly—using the wrong error terms to test statistical significance. The discussion of the issues surrounding the choice of experimental unit and analysis strategies goes beyond the scope of this article but should take place on the experimentation team.

Items 10 and 11 of Figure 2 are intended for the following three circumstances: First, when experimenters are statistically sophisticated and have a good idea of appropriate designs or analysis techniques;

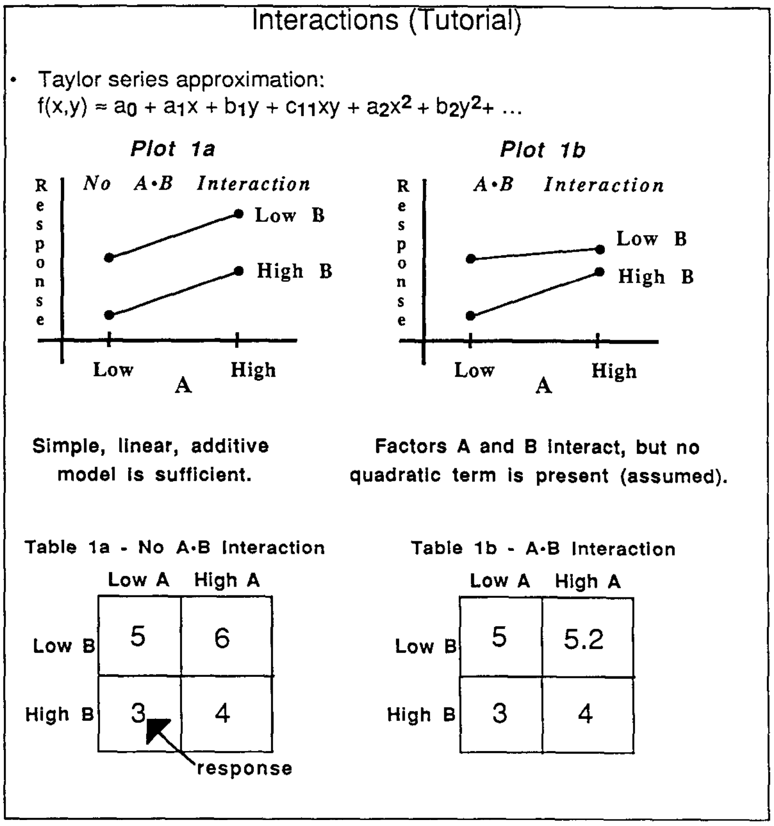

Figure 6. Graphical (tutorial) Presentation of Interactions.

second, when the experiment has been preceded by experiments in which a particular design or technique proved to be useful; third, when, on considering designs, analyses, and plots, the experimenters may want to change information in items 2–7—for example, narrowing the scope of the objective or increasing the number of settings for a control variable.

9. THE NEXT STAGE

By the time the experimentation team has come to a consensus concerning the information collected in items 1–10 of the guide sheet, the statistician (or surrogate) will have had the opportunity to step beyond the generic confines of the guide sheet and discuss more problem-specific issues that will affect the experimental design, such as multilevel factors, different sizes of experimental units, and logistics. Then, it may be useful to (a) choose candidate designs, (b) review them with the experimenters in the context of the collected information to determine if any of the designs should be dropped from further consideration, and (c) write an experimental design proposal that contains (at least) one or more proposed designs; a comparative analysis of the designs with respect to number of runs, resolution (or aliased effects), number of distinct control variable combinations, prediction error standard deviation, and so forth; a design recommendation with justification; and copies of the completed guide sheets.

When an experimental design has been selected, the sheets are used to help launch supportive tasks required for the experiment to be successful. This

Table 6. Interactions

Control variable	*y shift*	*z shift*	*Vendor*	*a shift*	*Speed*	*Height*	*Feed*
x shift			P				
y shift	—		P				
z shift	—	—	P				
Vendor	—	—	—	P			
a shift	—	—	—	—			
Speed	—	—	—	—	—		F, D
Height	—	—	—	—	—	—	

NOTE: Response variables are P = profile difference, F = surface finish, and D = surface defects.

involves issues addressed in items 11 and 12 on the guide sheet. Additionally, there will be logistical planning and planning for measurement capability studies, process capability studies, preexperiments to quantify the effects of various factors (held-constant and nuisance) on response variables, and trial runs.

In regard to item 12 of Figure 2, an experiment without a coordinator will probably fail. Though a statistician can play this role, it is often better played by another member of the experimentation team, who can "champion" the experiment among peers. The statistician (or surrogate) can play a strong support role and be primarily responsible for that in which he or she is professionally trained—the design and analysis of the experiment and *not* the execution.

Finally, considering item 13 of the guide sheet, the team should entertain the idea of trial runs to precede the experiment—especially if this is the first in a series of experiments. A trial run can consist of a centerpoint run or a small part (perhaps a block) of the experiment. The first and most important purpose of trial runs is to learn and refine experimental procedures without risking the loss of time and expensive experimental samples. Most experiments involve people (and sometimes machines) doing things that they have *never done before*. Usually some practice helps.

A second important reason for trial runs is to estimate experimental error before expending major resources. An unanticipated large experimental error could lead to canceling or redesigning the experiment, widening the ranges of settings, increasing the number of replicates, or refining the experimental procedure. An unanticipated small experimental error (does this ever really happen?) could have opposite effects on plans or cause the experimenters to reassess whether the estimate is right or complete.

A third reason is that trial runs are also excellent opportunities to ensure that data-acquisition systems are functioning and will permit experimental runs to be conducted as fast as had been planned.

Last, a fourth reason is that trial runs may yield results so unexpected that the experimenters decide to change their experimental plans.

Naturally, the feasibility and advisability of conducting trial runs depends on the context, but the experiment teams in which we have been involved have *never regretted* conducting trial runs. Some trial runs have saved experiments from disaster.

10. SUMMARY

To conduct complex experiments, careful planning with attention to detail is critical. Predesign planning is one part of the process by which experiments are conceived, planned, executed, and interpreted. It is often the part claimed by no one, hence it is often done informally—and sloppily. The use of predesign experiment guide sheets provides a way to systematize the process by which an experimentation team does this planning, to help people to (a) more clearly define the objectives and scope of an experiment and (b) gather information needed to design an experiment.

ACKNOWLEDGMENTS

We much appreciate the patient tolerance of the people who cooperated in the first use of this systematic approach, especially E. Malecki, R. Sanders, G. S. Smith, and R. Welsh, who provided the initial opportunity. G. Hahn, members of the Alcoa Laboratories Statistics group (L. Blazek, M. Emptage, A. Jaworski, K. Jensen, B. Novic, and D. Scott), P. Love, and M. Peretic also provided useful insight and comments. The thoughtful comments provided by the referees and editor considerably improved the article.

[*Received July 1991. Revised April 1992.*]

REFERENCES

Bishop, T., Petersen, B., and Trayser, D. (1982), "Another Look at the Statistician's Role in Experimental Planning and Design," *The American Statistician*, 36, 387–389.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: John Wiley.

Eisenhart, C. (1962), "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," *Journal of Research of the National Bureau of Standards*, 67C, 161–187.

Hahn, G. (1977), "Some Thing Engineers Should Know About Experimental Design," *Journal of Quality Technology*, 9, 13–20.

——— (1984), "Experimental Design in a Complex World," *Technometrics*, 26, 19–31.

Hoadley, A., and Kettenring, J. (1990), "Communications Between Statisticians and Engineers/Physical Scientists" (with commentary), *Technometrics*, 32, 243–274.

Hunter, W. G. (1977), "Some Ideas About Teaching Design of Experiments With 2^5 Examples of Experiments Conducted by Students," *The American Statistician*, 31, 12–17.

McCulloch, C. E., Boroto, D. R., Meeter, D., Polland, R., and Zahn, D. A. (1985), "An Expanded Approach to Educating Statistical Consultants," *The American Statistician*, 39, 159–167.

Montgomery, D. C. (1991), *Design and Analysis of Experiments* (3rd ed.) New York, John Wiley.

Natrella, M. G. (1979), "Design and Analysis of Experiments," in *Quality Control Handbook*, ed. J. M. Juran, New York: McGraw-Hill, pp. 27-35.