KEYS TO SUCCESSFUL DESIGNED EXPERIMENTS

Shari L. Kraber, CQE
Statistical Consultant
Stat-Ease Corporation
Minneapolis, MN 55413

SUMMARY

We can improve experimentation results by studying organizations that have experienced both frustrations due to poor experimentation methodology and satisfaction from successful applications. This paper identifies eight factors essential to successful experimentation. A solid understanding of these key factors is the foundation to a successful design of experiments program.

KEYWORDS

DOE, process understanding, continuous improvement

INTRODUCTION

Designed experiments are a powerful tool to achieve increased process understanding, leading to improvements that can result in significantly improved quality, decreased costs, and potentially thousands of dollars of savings for companies. So why don't more manufacturers use design of experiments (DOE)? Why do some companies try DOE, and then abandon it saying "it won't work for my company?" Using DOE successfully depends on understanding eight fundamental concepts. To illustrate these eight concepts, we'll follow a case study about experimentation done on an injection molding process.

1. DETERMINING THE OBJECTIVE OF THE EXPERIMENT.

The first decision before designing an experiment is "what is the objective, or purpose, of this study?" The focus of the study may be to screen out the factors that are not critical to the process, or it may be to optimize a few critical factors. A well-defined objective leads the experimenter to the correct DOE. The objective can also provide boundaries on the factor ranges or the number of levels to be studied.

For instance, if the experimenter is in the very initial stages of studying a process then the appropriate design choice may be a fractional two-level factorial. This DOE screens a large number of factors. It determines the vital few that are critical to the process. If the experimenter believes that the process is already close to optimum conditions, then a response surface design may be most appropriate. It will explore a few factors over many levels.

If you do not identify the objectives of a study, you may pay the consequences:

* trying to study too many or too few factors
* not completing replications as needed
* not measuring the correct responses
* arriving at conclusions that are already known

In essence, vague objectives lead to lost time and money, as well as feelings of frustration for all involved. Identifying the objective up front builds a common understanding and mutually agreed upon expectations of the study.

In our case study of the injection molder, management wants to reduce variation in the shrinkage of their parts. If the shrinkage can be stabilized, then mold dimensions can be adjusted so the parts can be made consistently. The objective is: Identify factor levels in the injection molding process that control shrinkage. The factors and levels to be studied are:

FACTOR NAME	UNITS	LOW LEVEL	HIGH LEVEL
A: Mold Temp	deg F	130	180
B: Holding Pressure	psig	1200	1500
C: Booster Pressure	psig	1500	1800
D: Moisture	%	.05	.15
E: Screw Speed	inches/sec	1.5	4.0
F: Cycle Time	sec	25	30
G: Gate Size	thousands	30	50

2. MEASURING YOUR RESPONSE.

Many DOE's fail because the response cannot be measured quantitatively. A classic example is found with visual inspections for quality. Traditionally, process operators or inspectors have developed a qualitative system that they use to determine if a product passes or fails. At best, they may have boundary samples of minimally acceptable product. Although this system may be OK for production, it does not have enough precision for a good DOE. Pass/fail data can be used in DOE, but it's very crude. For example, if your process typically produces a 0.1% defective rate you would expect to find 5 out of 5000 parts defective. In order to execute a simple designed experiment that investigated 3 factors in 8 experimental runs on such a process, you would need to utilize a minimum of 40,000 parts (8 x 5000). This would assure getting enough defects to judge improvement, but can you afford the cost?

For the purposes of experimentation, a rating system works much better. Select a scale, perhaps 1 (worst) to 10 (best), that each unit can be rated against. Define the scale as well as you can by providing descriptions of what the unit would look like at several points on the scale. Train three to five people to use the scale. During the experiment, each trained inspector should rate each unit. Then you can use the average rating from all the inspectors as the response. Even crude scaling from 1 to 5 will be far better than the simple pass/fail method.

Another aspect to consider is whether or not the measurement is repeatable. Can the measurement be performed on the same unit multiple times and produce the same results? If a measurement exhibits lots of variability, this may overshadow the effects that you are looking for. During the initial training of inspectors, you should emphasize the importance of minimizing inspector variability in order to reduce reproducibility errors. For a good DOE, the testing method must consistently produce reliable results.

In the case study on injection molding, the experimenters will measure percent shrinkage at a critical dimension on the part.

3. REPLICATION - UNDERSTANDING YOUR SIGNAL TO NOISE RATIO.

Replication is a tool used to improve the chance of detecting a statistically significant effect in the midst of natural process variation. In some processes, the amount of normal process variation is so large that it easily obscures the factor effects that you are trying to detect. Designed experiments use averaging to

increase statistical power, the probability of finding a statistically significant result. The question remains – how many runs are needed to identify significant effects, given the current process variation? Signal to noise ratios help you determine the minimal experimental runs needed to achieve a given power, or statistical significance, for your DOE. The signal is the change in your response that you want to detect. You have to determine the smallest change that you want to be able to detect. The noise is the random variation that occurs in the response during normal operation. The noise in the response can be estimated in the following ways:

- from repeatability studies on the process
- control charts (R-bar divided by d_2)
- the root mean square error from a designed experiment ANOVA report
- a best guess estimate

You can use the following table to determine how many two-level factorial experiments you need to run for the desired detectable signal. If the minimum runs exceeds what you can afford, then you must find a way to decrease noise (variability caused by the process, sampling and/or testing) or accept an increase in the minimum detectable signal (the minimum effect that will be detected by the experiment).

Signal to Noise Ratio (Δ/σ)	Minimum Number of Runs
1.0	**64**
1.4	**32**
2.0	**16**
2.8	**8**

The derivation is based on providing approximately a 90% probability of finding an effect (signal) of size delta (Δ). The number of runs, rounded up to the nearest power of two, comes from the formula: $N = (4r)^2 / (\Delta/\sigma)^2 = 64 / (\Delta/\sigma)^2$ where N is the total number of experiments, r is number of levels of the factors (two in this case), Δ is the signal to detect, and σ is the noise level, or standard deviation of a single experiment. There is a 90% chance of revealing an active effect of size Δ, and a 5% chance of falsely revealing an inactive effect as active. (E.G. Watts, 1997) (R.E. Wheeler, 1979)

Don't mistake repeated measurements with replication. Suppose an experiment is set up with two replicates of 8 runs each. The 8 runs are made and two units are taken from each run and sent to the lab for testing. This is an example of a repeated measurement. The test results from the two units may be different, but they come from the same experimental run, so the difference only represents variability during that run. True replication of an experiment must consist of going through all the setup procedures. Operators and managers typically resist this type of replication because it seems to be wasted effort. However, true replication will allow the data to be an accurate representation of all the variability in the process. It will yield better results in the long run. Do it right the first time, or you'll just have to do it over later!

For our case study, from previous experimentation on the injection molding equipment, the standard deviation of the process is known to be 0.60. Management would like to detect an effect of magnitude 0.85. Therefore the signal to noise ratio is approximately 1.4. The appropriate number of runs for this two-level factorial experiment is 32 runs.

4. RANDOMIZATION.

Experiments set up by hand generally follow a very structured sequence, sometimes referred to as standard design order. The order in which you run the experiments should be randomized to avoid influence by time-related, uncontrolled, variables. Tool wear, temperature, humidity, and raw materials are examples of variables that could change during an experiment without your being aware of it. Without randomization of the experimental runs, these changes could influence the response and be incorrectly attributed to a factor effect. For example, assume that a hypothetical experiment is performed so that all the low levels of Factor A are run first, followed by all the high levels of Factor A. During the course of the experiment, the humidity in the area changes by 50%, creating a significant effect on the response. In the analysis stage, Factor A then appears to be significant, but it really is the change in the humidity level that caused the effect. Randomization would have prevented this confusion. Always do it!

5. BLOCKING TO REDUCE VARIATION.

Blocking is a tool used to separate some expected variation from the analysis of the factor effects. For example, you may need to run an experiment over the course of several days. Each day the process may vary slightly. By placing the runs from each day into a separate block, the day to day variation can be eliminated from the analysis and make it easier to detect significant effects. Blocking screens out noise caused by known sources of variation, such as raw material batch, shift changes, or machine differences. It divides the experimental runs into homogeneous groups and then arithmetically removes the difference. Removing this block effect reduces the noise in the experiment, increasing the sensitivity of your DOE.

Do not block on anything that you want to study. For example, if you want to measure the difference between two raw material suppliers include them as a factor to study in your DOE. Although blocking is a very effective tool for removing expected variation you should be aware that it is not without a cost. Blocking slightly reduces the amount of information that can be obtained because the blocking factor will be aliased with another factor. Aliasing is discussed in the next section. Usually the benefits of blocking outweigh the slight loss of information.

In the injection molding case study, management wants to use four identical injection-molding machines for the runs. Because there may be slight differences between the machines, the experiment is blocked on machine. Therefore each machine will have eight experimental runs to complete. In this case, the cost of blocking is that the interaction of A: Mold Temperature and B: Holding Pressure cannot be estimated. A nice side benefit of this blocking scheme – the experiment will get done 4 times faster!

6. UNDERSTANDING THE EXISTENCE OF THE ALIAS STRUCTURE.

The alias structure is a critical and often over-looked feature of standard two-level factorial, Plackett-Burman, Taguchi designs, or any other design that's done in fractional form. What is an alias? In the more common sense, it is a name taken on by a person as a type of disguise. In designed experiments, if a fractional factorial design is run there are not enough individual pieces of information to independently estimate all the effects. So, the effects combine together and become aliases for each other. For example, if you try to study 4 factors in only 8 runs, a half-fraction, each main effect is aliased with a three-factor interaction, and each two-factor interaction is aliased with another two-factor interaction. Counting the mean we can estimate 8 pieces of information from the 8 runs - no more.

What does this mean to the average experimenter? Be careful when analyzing the results of fractional factorial, Plackett-Burman, or Taguchi designs. These designs may indicate that a main effect or a two-factor interaction is significant, when in reality it could just as easily be any one of the aliases. An understanding of alias structures can be a tremendous asset to the experimenter. Software packages like Design-Expert from Stat-Ease Corporation allow the user to easily view the alias structure of a design

before assigning factor names. For more explanation of alias structures, see *Statistics for Experimenters* by Box, Hunter and Hunter.

In the injection molding case study there are seven factors to be studied in 32 out of a possible 128 runs, so it's a ¼ fractional factorial design. This is a Resolution IV design, which means that main effects may be aliased with three factor interactions, and two factor interactions may be aliased with each other. In this case, the alias structure shows the experimenters that CE = FG, CF = EG, and CG = EF. The experimenter will assign to C, E, F and G the factors least likely to interact. If interactions do occur, care will have to be taken during the analysis so that the results will be interpreted properly.

7. ITERATIVE EXPERIMENTATION

Designed experiments should be executed in an iterative manner so that information that is learned in one experiment can be applied in the next experiment. For example, rather than running a very large experiment with many factors and using up the majority of your resources, consider starting with a smaller experiment and then building upon the results. A typical series of experiments consists of a screening design (fractional factorial) to identify the significant factors, a full factorial or response surface design to fully characterize or model the effects, followed up with confirmation runs to verify your results. If you make a mistake in the selection of your factor ranges or responses in a very large experiment, it can be very costly. A series of iterative experiments allows you to reassess your selections and make changes midstream at a relatively low cost.

In the injection molding case, previous experimentation determined that these seven factors can influence shrinkage. It is expected that follow-up studies or confirmation runs will need to be completed.

8. CONFIRMATION OF RESULTS.

After all the effort that goes into planning, running, and analyzing a designed experiment, it is very exciting to get the results of your work. There is a tendency to eagerly grab the results and rush out to production and say, "We have the answer! This will solve the problem!" BEFORE doing that, you need to take the time to do a confirmation run and verify the results of your experiment! Some software packages will provide you with a confidence interval with which you can compare your confirmation run results. You may want to look at both a prediction interval for a single point and a confidence interval for a mean, or an average of several runs. Remember that in statistics you never deal with absolutes - there is always uncertainty in your recommendations. Be sure to double-check your results.

In the injection molding case, the results of the experiment showed that the significant factors were interactions between Mold Temperature and Moisture Content, Booster Pressure and Moisture Content, and Holding Pressure and Cycle Time. Shrinkage could be stabilized by using a low moisture level in the pellets. At this condition the setting of the mold temperature and the booster pressure didn't matter. For economic considerations, the low cycle time was paired with the high holding pressure. These results were confirmed through subsequent confirmation studies.

CONCLUSION

Design of experiments is a very effective tool that can be utilized in virtually all industries. To effectively use this tool, you need to learn and practice the proper techniques. We present eight keys to success:

- Determining the Objective of the Experiment
- Measuring your Response

- Replication - Understanding Your Signal to Noise Ratio
- Randomization
- Blocking to Reduce Variation
- Understanding the Existence of the Alias Structure
- Iterative Experimentation
- Confirming your Results

Practice these fundamental concepts to effectively utilize design of experiments in your company.

REFERENCE LIST

Box, Hunter and Hunter 1978. *Statistics For Experimenters*. John Wiley & Sons.
Helseth, T.H. 1997. *Design-Expert for Windows v5*. Stat-Ease, Inc.
Watts, E. G. 1997. Explaining Power Using Two-Level Factorial Designs. *Journal of Quality Technology*, Vol. 29, No. 3, (July).
Wheeler, R. E. 1979. Portable Power. *Technometrics* Vol. 16, No. 2, (May).