THE STARTING POINT FOR IMPROVEMENT IS TO RECOGNIZE THE NEED.

IMAI

Improve

The Improve topic is presented in the following topic areas:

- Design of experiments (DOE)
- Lean methods
- Implementation

Design of Experiments

Design of Experiments is discussed in the following topic areas:

- Introduction
- Terminology
- Design principles
- Planning experiments
- One factor designs
- Fractional factorial designs
- Full factorial designs

DOE Introduction*

Classical experiments focus on 1FAT (one factor at a time) at two or three levels and attempt to hold everything else constant (which is impossible to do in a complicated process). When DOE is properly constructed, it can focus on a wide range of key input factors or variables and will determine the optimum levels of each of the factors. It should be recognized that the Pareto principle applies to the world of experimentation. That is, 20% of the potential input factors generally make 80% of the impact on the result.

The classical approach to experimentation, changing just one factor at a time, has shortcomings:

- Too many experiments are necessary to study the effects of all the input factors.
- The optimum combination of all the variables may never be revealed.

* A substantial portion of the material throughout Section IX comes from the *CQE Primer* by Wortman (2012)[25].

DOE Introduction (Continued)

- The interaction (the behavior of one factor may be dependent on the level of another factor) between factors cannot be determined.
- Unless carefully planned and the results studied statistically, conclusions may be wrong or misleading.
- Even if the answers are not actually wrong, non-statistical experiments are often inconclusive. Many of the observed effects tend to be mysterious or unexplainable.
- Time and effort may be wasted through studying the wrong variables or obtaining too much or too little data.

Design of experiments overcomes these problems by careful planning. In short, DOE is a methodology of varying a number of input factors simultaneously, in a carefully planned manner, such that their individual and combined effects on the output can be identified. Advantages of DOE include:

- Many factors can be evaluated simultaneously, making the DOE process economical and less interruptive to normal operations.
- Sometimes factors having an important influence on the output cannot be controlled (noise factors), but other input factors can be controlled to make the output insensitive to noise factors.
- In-depth, statistical knowledge is not always necessary to get a big benefit from standard planned experimentation.
- One can look at a process with relatively few experiments. The important factors can be distinguished from the less important ones. Concentrated effort can then be directed at the important ones.
- Since the designs are balanced, there is confidence in the conclusions drawn. The factors can usually be set at the optimum levels for verification.
- If important factors are overlooked in an experiment, the results will indicate that they were overlooked.
- Precise statistical analysis can be run using standard computer programs.
- Frequently, results can be improved without additional costs (other than the costs associated with the trials). In many cases, tremendous cost savings can be achieved.

DOE Terms

Alias — An alias occurs when two factor effects are confused or confounded with each other. See confounded.

Balanced design — A fractional factorial design, in which an equal number of trials (at every level state) is conducted for each factor.

Block — A subdivision of the experiment into relatively homogenous experimental units. The term is from agriculture, where a single field would be divided into blocks for different treatments.

Blocking — When structuring fractional factorial experimental test trials, blocking is used to account for variables that the experimenter wishes to avoid. A block may be a dummy factor which doesn't interact with the real factors.

Box-Behnken — When full, second-order, polynomial models are to be used in response surface studies of three or more factors, Box-Behnken designs are often very efficient. They are highly fractional, three-level factorial designs.

Collinear — A collinear condition occurs when two variables are totally correlated. One variable must be eliminated from the analysis for valid results.

Confounded — When the effects of two factors are not separable. In the following example, A, B, and C are input factors and columns AB, AC, & BC represent interactions (multiplication of 2 factors). Confounded columns are identified by arrows, indicating the setting of one cannot be separated from the setting of the other. See the two examples below:

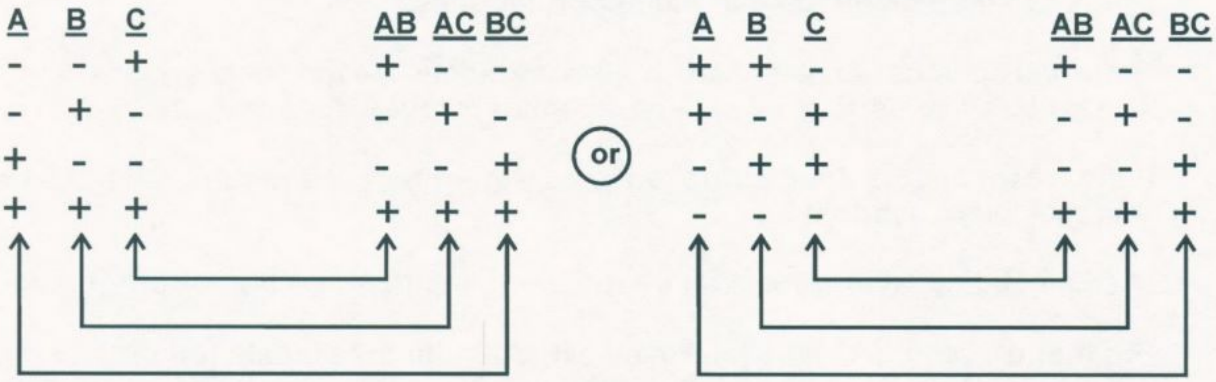


Figure 9.1 A is confounded with BC, B with AC, and C with AB

DOE Terms (Continued)

Correlation coefficient (r)
A number between -1 and 1 that indicates the degree of linear relationship between two sets of numbers. Zero (0) indicates no linear relationship.

Covariates
Things which change during an experiment which had not been planned to change, such as temperature or humidity. Randomize the test order to alleviate this problem. Record the value of the covariate for possible use in regression analysis.

Curvature
Refers to non-straight line behavior between one or more factors and the response. Curvature is usually expressed in mathematical terms involving the square or cube of the factor. For example, in the model:

$$Y = B_0 + B_1X_1 + B_{11}(X_1 \cdot X_1) + \varepsilon$$

The term $B_{11}(X_1 \cdot X_1)$ describes curvature.

Degrees of freedom
The terms used are DOF, DF, df or ν. The number of measurements that are independently available for estimating a population parameter.

Design of experiments (DOE)
The arrangement in which an experimental program is to be conducted and the selection of the levels of one or more factors or factor combinations to be included in the experiment. Factor levels are accessed in a balanced full or fractional factorial design. The term SDE (statistical design of experiment) is also widely used.

Efficiency
A concept from R. A. Fisher. He considered one estimator more efficient than another if it had a smaller variance. Percentage efficiency is calculated as:

$$\frac{\text{variance of minimum estimator}}{\text{variance of an estimator in question}} \times 100$$

EVOP
Stands for evolutionary operation, a term that describes the way sequential experimental designs can be made to adapt to system behavior by learning from present results and predicting future treatments for better response. Often, small response improvements may be made via large sample sizes. The experimental risk, however, is quite low because the trials are conducted in the near vicinity of an already satisfactory process.

DOE Terms (Continued)

Experiment — A test undertaken to make an improvement in a process or to learn previously unknown information.

Experimental error — Variation in response or outcome of virtually identical test conditions. This is also called residual error.

First-order — Refers to the power to which a factor appears in a model. If "X_1" represents a factor and "B" is its factor effect, then the model:

$$Y = B_0 + B_1X_1 + B_2X_2 + \varepsilon$$

is first-order in both X_1 and X_2. First-order models cannot account for curvature or interaction.

Fractional — An adjective that means fewer experiments than the full design calls for. The three-factor designs shown below are two-level, half-fractional designs.

A	B	C
-	-	-
-	+	+
+	-	+
+	+	-

A	B	C
-	-	+
-	+	-
+	-	-
+	+	+

Full factorial — Describes experimental designs which contain all combinations of all levels of all factors. No possible treatment combinations are omitted. A two-level, three-factor full factorial design is shown below:

A	B	C
-	-	-
-	-	+
-	+	-
-	+	+
+	-	-
+	-	+
+	+	-
+	+	+

Input factor — An independent variable which may affect a (dependent) response variable and is included at different levels in the experiment.

DOE Terms (Continued)

Inner array In Taguchi-style fractional factorial experiments, these are the factors that can be controlled in a process.

Interaction An interaction occurs when the effect of one input factor on the output depends upon the level of another input factor. Refer to the following diagrams:

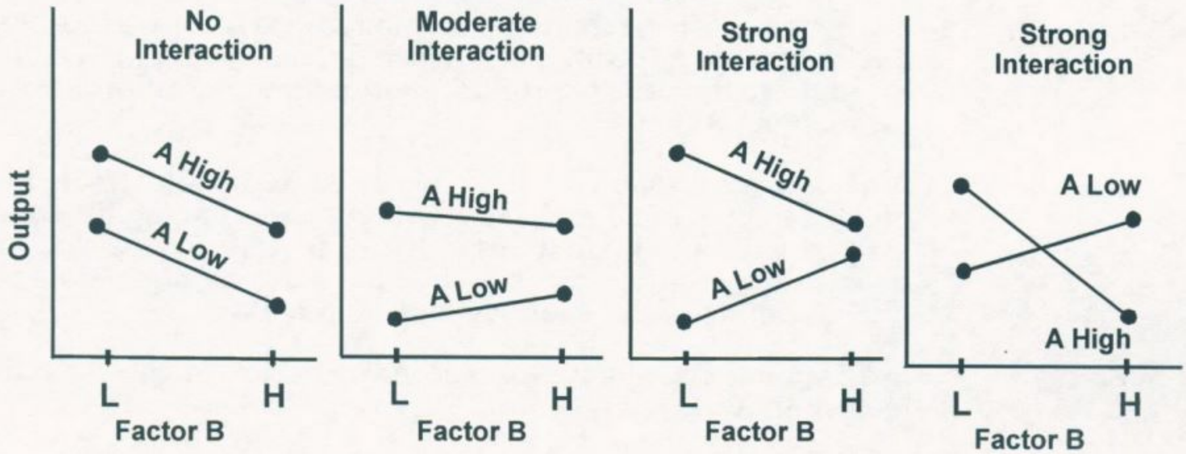


Figure 9.2 Examples of Interactions

Interactions can be readily examined with full factorial experiments. Often, interactions are lost with fractional factorial experiments.

Level A given factor or a specific setting of an input factor. Four levels of a heat treatment may be 100°F, 120°F, 140°F, and 160°F.

Main effect An estimate of the effect of a factor independent of any other factors.

Mixture experiments Experiments in which the variables are expressed as proportions of the whole and sum to 1.0.

Multi-collinearity This occurs when two or more input factors are expected to independently affect the value of an output factor, but are found to be highly correlated. For example, an experiment is being conducted to determine the market value of a house. The input factors are square feet of living space and number of bedrooms. In this case, the two input factors are highly correlated. Larger residences have more bedrooms.

DOE Terms (Continued)

Nested experiments	An experimental design in which all trials are not fully randomized. There is generally a logical reason for taking this action. For example, in an experiment, technicians might be nested within labs. As long as each technician stays with the same lab, the techs are nested. It is not often that techs travel to different labs just to make the design balanced.
Optimization	Involves finding the treatment combinations that give the most desired response. Optimization can be "maximization" (as, for example, in the case of product yield) or "minimization" (in the case of impurities).
Orthogonal	A design is orthogonal if the main and interaction effects in a given design can be estimated without confounding the other main effects or interactions. A full factorial is said to be balanced, or orthogonal, because there are an equal number of data points under each level of each factor.
Outer array	In a Taguchi-style fractional factorial experiment, these are the factors that cannot be controlled in a process.
Paired comparison	The basis of a technique for treating data so as to ignore sample-to-sample variability and focus more clearly on variability caused by a specific factor effect. Only the differences in response for each sample are tested because sample-to-sample differences are irrelevant.
Parallel experiments	These experiments are done at the same time, not one after another, e.g., agricultural experiments in a big corn field. Parallel experimentation is the opposite of sequential experimentation.
Precision	The closeness of agreement between test results.
Qualitative	This refers to descriptors of category and/or order, but not of interval or origin. Different machines, operators, materials, etc. represent qualitative levels or treatments.
Quantitative	This refers to descriptors of order and interval (interval scale) and possibly also of origin (ratio scale). As a quantitative factor, "temperature" might describe the interval value 27.32°C. As a quantitative response, "yield" might describe the ratio value 87.42%.
Randomized trials	Frees an experiment from the environment and eliminates biases. This technique avoids the undue influences of systematic changes that are known or unknown.

DOE Terms (Continued)

Repeated trials	Trials that are conducted to estimate the pure trial-to-trial experimental error so that lack of fit may be judged. Also called replications.
Residual error (ε) or (E)	The difference between the observed and the predicted value for that result, based on an empirically determined model. It can be variation in outcomes of virtually identical test conditions.
Residuals	The difference between experimental responses and predicted model values.
Resolution I	An experiment in which tests are conducted, adjusting one factor at a time, hoping for the best. This experiment is not statistically sound (definition totally fabricated by the authors).
Resolution II	An experiment in which some of the main effects are confounded. This is very undesirable.
Resolution III	A fractional factorial design in which no main effects are confounded with each other, but the main effects and two factor interaction effects are confounded.
Resolution IV	A fractional factorial design in which the main effects and two factor interaction effects are not confounded, but the two factor effects may be confounded with each other.
Resolution V	A fractional factorial design in which no confounding of main effects and two factor interactions occur. However, two factor interactions may be confounded with three factor and higher interactions.
Resolution VI	Also called Resolution V+. This is at least a full factorial experiment with no confounding. It can also mean two blocks of 16 runs.
Resolution VII	Can refer to eight blocks of 8 runs.
Response surface methodology (RSM)	The graph of a system response plotted against one or more system factors. Response surface methodology employs experimental design to discover the "shape" of the response surface and then uses geometric concepts to take advantage of the relationships discovered.
Response variable	The variable that shows the observed results of an experimental treatment. Also known as the output or dependent variable.

DOE Terms (Continued)

Robust design — A term associated with the application of Taguchi experimentation in which a response variable is considered robust or immune to input variables that may be difficult or impossible to control.

Screening experiment — A technique to discover the most (probable) important factors in an experimental system. Most screening experiments employ two-level designs. A word of caution about the results of screening experiments: if a factor is not highly significant, it does not necessarily mean that it is insignificant.

Second-order — Refers to the power to which one or more factors appear in a model. If "X_1" represents a factor and "B_1" is its factor effect, then the model:

$$Y = B_0 + B_1X_1 + B_{11}(X_1 \cdot X_1) + B_2X_2 + \varepsilon$$

is second-order in X_1 but not in X_2. Second-order models can account for curvature and interaction. $B_{12}(X_1 \cdot X_2)$ is another second-order example, representing an interaction between X_1 and X_2.

Sequential experiments — Experiments are done one after another, not at the same time. This is often required by the type of experimental design being used. Sequential experimentation is the opposite of parallel experimentation.

Simplex — A geometric figure that has a number of vertexes (corners) equal to one more than the number of dimensions in the factor space.

Simplex design — A spatial design used to determine the most desirable variable combination (proportions) in a mixture.

Test coverage — The percentage of all possible combinations of input factors in an experimental test.

Treatments — In an experiment, the various factor levels that describe how an experiment is to be carried out. A pH level of 3 and a temperature level of 37° Celsius describe an experimental treatment.

Applications of DOE*

Situations, where experimental design can be effectively used include:

- Choosing between alternatives

- Selecting the key factors affecting a response

- Response surface modeling to:

  - Hit a target
  - Reduce variability
  - Maximize or minimize a response
  - Make a process robust (despite uncontrollable "noise" factors)
  - Seek multiple goals

DOE Steps*

Getting good results from a DOE involves a number of steps:

- Set objectives
- Select process variables
- Select an experimental design
- Execute the design
- Check that the data are consistent with the experimental assumptions
- Analyze and interpret the results
- Use/present the results (may lead to further runs or DOEs)

Important practical considerations in planning and running experiments are:

- Check the performance of gauges/measurement devices first
- Keep the experiment as simple as possible
- Check that all planned runs are feasible
- Watch out for process drifts and shifts during the run
- Avoid unplanned changes (e.g. switching operators at half time)
- Allow some time (and back-up material) for unexpected events
- Obtain buy-in from all parties involved
- Maintain effective ownership of each step in the experimental plan
- Preserve all the raw data - do not keep only summary averages!
- Record everything that happens
- Reset equipment to its original state after the experiment

* Modified from NIST (2001)[12] and used with permission.

Experimental Objectives*

Choosing an experimental design depends on the objectives of the experiment and the number of factors to be investigated. Some experimental design objectives are discussed below:

1. Comparative objective: If several factors are under investigation, but the primary goal of the experiment is to make a conclusion about whether a factor, in spite of the existence of the other factors, is "significant," then the experimenter has a comparative problem and needs a comparative design solution.

2. Screening objective: The primary purpose of this experiment is to select or screen out the few important main effects from the many lesser important ones. These screening designs are also termed main effects or fractional factorial designs.

3. Response surface (method) objective: This experiment is designed to let an experimenter estimate interaction (and quadratic) effects and, therefore, give an idea of the (local) shape of the response surface under investigation. For this reason, they are termed response surface method (RSM) designs. RSM designs are used to:

    - Find improved or optimal process settings
    - Troubleshoot process problems and weak points
    - Make a product or process more robust against external influences

4. Optimizing responses when factors are proportions of a mixture objective: If an experimenter has factors that are proportions of a mixture and wants to know the "best" proportions of the factors to maximize (or minimize) a response, then a mixture design is required.

5. Optimal fitting of a regression model objective: If an experimenter wants to model a response as a mathematical function (either known or empirical) of a few continuous factors, to obtain "good" model parameter estimates, then a regression design is necessary.

Response surface, mixture, and regression designs are not featured as separate entities in this Primer. It should be noted that most good computer programs will provide these models. The best design sources are often full factorial (in some cases with replication) and screening designs. A summary of the first 3 objectives above are outlined in Table 9.3.

* Modified from NIST (2001)[12] and used with permission.

Select and Scale the Process Variables

Process variables include both inputs and outputs, I. e. factors and responses. The selection of these variables is best done as a team effort. The team should:

- Include all important factors (based on engineering and operator judgments)
- Be bold, but not foolish, in choosing the low and high factor levels
- Avoid factor settings for impractical or impossible combinations
- Include all relevant responses
- Avoid using responses that combine two or more process measurements

When choosing the range of settings for input factors, it is wise to avoid extreme values. In some cases, extreme values will give runs that are not feasible; in other cases, extreme ranges might move the response surface into some erratic region.

The most popular experimental designs are called two-level designs. Two-level designs are simple and economical and give most of the information required to go to a multi-level response surface experiment, if one is needed. However, two-level designs are something of a misnomer. It is often desirable to include some center points (for quantitative factors) during the experiment (center points are located in the middle of the design "box.").

Design Guidelines

Number of Factors	Comparative Objective	Screening Objective	Response Surface Objective
1	1-factor completely randomized design	—	—
2 - 4	Randomized block design	Full or fractional factorial	Central composite or Box-Behnken
5 or more	Randomized block design	Fractional factorial or Plackett-Burman	Screen first to reduce number of factors

Table 9.3 Design Selection Guidelines

The choice of a design depends on the amount of resources available and the degree of control over making wrong decisions (Type I and Type II hypotheses errors). It is a good idea to choose a design that requires somewhat fewer runs than the budget permits, so that additional runs can be added to check for curvature and to correct any experimental mishaps.

A Typical DOE Checklist

Every experimental investigation will differ in detail, but the following checklist will be helpful for many investigations.

- Define the objective of the experiment.

- The principle experimenter should learn as many facts about the process as possible prior to brainstorming.

- Brainstorm a list of the key independent and dependent variables with people knowledgeable of the process and determine if these factors can be controlled or measured.

- Run "dabbling experiments" where necessary to debug equipment or determine measurement capability. Develop experimental skills and get some preliminary results.

- Assign levels to each independent variable in the light of all available knowledge.

- Select a standard DOE plan or develop one by consultation. It pays to have one person outline the DOE and another review it critically.

- Run the experiments in random order and analyze results periodically.

- Draw conclusions. Verify by replicating experiments, if necessary, and proceed to follow-up with further experimentation if an improvement trend is indicated in one or more of the factors.

It is often a mistake to believe that "one big experiment will give the answer." A more useful approach is to recognize that while one experiment might give a useful result, it is more common to perform two, three, or more experiments before a complete answer is attained. An iterative approach is usually the most economical. Putting all one's eggs in one basket is not advisable. It is logical to move through stages of experimentation, each stage supplying a different kind of answer.

Experimental Assumptions*

In all experimentation, one makes assumptions. Some of the engineering and mathematical assumptions an experimenter can make include:

- Are the measurement systems capable for all responses?
- Is the process stable?
- Are the residuals (the difference between the model predictions and the actual observations) well behaved?

Is the Measurement System Capable?

It is not a good idea to find, after finishing an experiment, that the measurement devices are incapable. This should be confirmed before embarking on the experiment itself. In addition, it is advisable, especially if the experiment lasts over a protracted period, that a check be made on all measurement devices from the start to the conclusion of the experiment. Strange experimental outcomes can often be traced to 'hiccups' in the metrology system.

Is the Process Stable?

Experimental runs should have control runs which are done at the "standard" process set points, or at least at some identifiable operating conditions. The experiment should start and end with such runs. A plot of the outcomes of these control runs will indicate if the underlying process itself drifted or shifted during the experiment. It is desirable to experiment on a stable process. However, if this cannot be achieved, then the process instability must be accounted for in the analysis of the experiment.

Are the Residuals Well Behaved?

Residuals are estimates of experimental error obtained by subtracting the observed response from the predicted response. The predicted response is calculated from the chosen model, after all the unknown model parameters have been estimated from the experimental data.

* Modified from NIST (2001)[12] and used with permission.

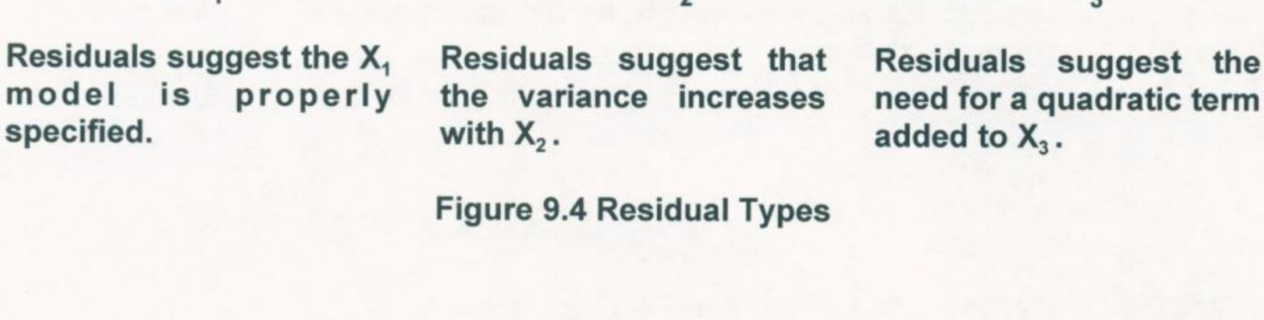
Experimental Assumptions (Continued)

Are the Residuals Well Behaved? (Continued)

Residuals can be thought of as elements of variation unexplained by the fitted model. Since this is a form of error, the same general assumptions apply to the group of residuals that one typically uses for errors in general: one expects them to be normally and independently distributed with a mean of 0 and some constant variance.

These are the assumptions behind ANOVA and classical regression analysis. This means that an analyst should expect a regression model to err in predicting a response in a random fashion; the model should predict values higher and lower than actual, with equal probability. In addition, the level of the error should be independent of when the observation occurred in the study, or the size of the observation being predicted, or even the factor settings involved in making the prediction.

The overall pattern of the residuals should be similar to the bell-shaped pattern observed when plotting a histogram of normally distributed data. Graphical methods are used to examine residuals. Departures from assumptions usually mean that the residuals contain structure that is not accounted for in the model. Identifying that structure, and adding a term representing it to the original model, leads to a better model. Any graph suitable for displaying the distribution of a set of data is suitable for judging the normality of the distribution of a group of residuals. The three most common types are: histograms, normal probability plots, and dot plots. Shown below are examples of dot plot results.



Residuals suggest the X_1 model is properly specified.

Residuals suggest that the variance increases with X_2.

Residuals suggest the need for a quadratic term added to X_3.

Figure 9.4 Residual Types

Interaction Case Study

A simple 2 x 2 factorial experiment (with replication) was conducted in the textile industry. The response variable was ED/MSH (ends down/thousand spindle hours.) The independent factors were RH (relative humidity) and ion level (the environmental level of negative ions). Both of these factors were controllable. A low ED/MSH is desirable since fewer thread breaks means higher productivity.

An ANOVA showed the main effects were not significant but the interaction effects were highly significant. Consider the data table and plots in Figure 9.5:

Main Effects	
Factor Level	ED/MSH Average
RH 37%	7.75
RH 41%	7.85
Low Ion	7.9
High Ion	7.7

Interaction Effects		
	RH 37%	RH 41%
Low Ion	7.2	8.3
High Ion	8.6	7.1

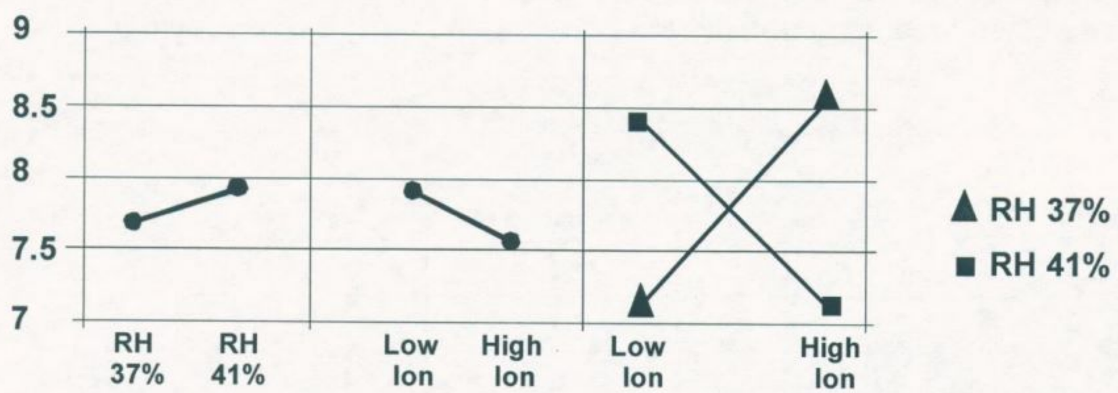


Figure 9.5 Relative Humidity and Ion Level Effects on ED/MSH

The above interaction plot demonstrates that if the goal is to reduce breaks, an economic choice could be made between low ion/low RH and high ion/high RH.

Randomized Block Plans

In comparing a number of factor treatments, it is desirable that all other conditions be kept as nearly constant as possible. The required number of tests may be too large to be carried out under similar conditions. In such cases, one may be able to divide the experiment into blocks, or planned homogeneous groups. When each group in the experiment contains exactly one measurement of every treatment, the experimental plan is called a randomized block plan. A randomized block design for air permeability response is shown below:

	Fabric Types I, II, III, & IV			
	I	II	III	IV
Chemical Applications A, B, C, D	B(15.1)	D(11.6)	A(15.4)	C(9.9)
	C(12.2)	C(13.1)	B(16.3)	D(9.4)
	A(19.0)	B(17.6)	D(16.0)	B(8.6)
	D(11.5)	A(13.0)	C(10.8)	A(11.5)

An experimental scheme may take several days to complete. If one expects some biasing differences among days, one might plan to measure each item on each day, or to conduct one test per day on each item. A day would then represent a block. A randomized incomplete block (tension response) design is shown below.

	Treatment			
Block (Days)	A	B	C	D
1	-5	Omitted	-18	-10
2	Omitted	-27	-14	-5
3	-4	-14	-23	Omitted
4	-1	-22	Omitted	-12

Only treatments A, C, and D are run on the first day. B, C, and D on the second, etc. In the whole experiment, note that each pair of treatments, such as BC, occur twice together. The order in which the three treatments are run on a given day follows a randomized sequence.

Blocking factors are commonly environmental phenomena outside of the control of the experimenter.

Latin Square Designs

A Latin square design is called a one-factor design because it attempts to measure the effects of a single key input factor on an output factor. The experiment further attempts to block (or average) the effects of two or more nuisance factors. Such designs were originally applied in agriculture when the two sources of non-homogeneity (nuisance factors) were the two directions on the field. The square was literally a plot of ground.

In Latin square designs, a third variable, the experimental treatment, is then applied to the source variables in a balanced fashion. The Latin square plan is restricted by two conditions:

- The number of rows, columns, and treatments must be the same.
- There should be no expected interactions between row and column factors, since these cannot be measured. If there are, the sensitivity of the experiment is reduced.

A Latin square design is essentially a fractional factorial experiment which requires less experimentation to determine the main treatment results.

Consider the following 5 x 5 Latin square design:

	Carburetor Type				
Driver	I	II	III	IV	V
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

In the above design, five drivers and five carburetors were used to evaluate gas mileage from five cars (A, B, C, D, and E). Note that only twenty-five of the potential 125 combinations are tested. Thus, the resultant experiment is a one-fifth fractional factorial. Similar 3 x 3, 4 x 4, and 6 x 6 designs may be utilized.

In some situations, what is thought to be a nuisance factor can end up being very important.

Graeco-Latin Designs

Graeco-Latin square designs are sometimes useful to eliminate more than two sources of variability in an experiment. A Graeco-Latin design is an extension of the Latin square design, but one extra blocking variable is added for a total of three blocking variables.

Consider the following 4 X 4 Graeco-Latin design:

	Carburetor Type			
Driver	**I**	**II**	**III**	**IV**
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

Cars
A,B,C,D

Days
α,β,γ,δ

The output (response) variable could be gas mileage for the 4 cars (A, B, C, D).

Hyper-Graeco-Latin Designs

A hyper-Graeco-Latin square design permits the study of treatments with more than three blocking variables.

Consider the following 4 x 4 hyper-Graeco-Latin design:

	Carburetor Type			
Driver	**I**	**II**	**III**	**IV**
1	AαMφ	BβNX	CγOΨ	DδPΩ
2	BδNΩ	AγMΨ	DβPX	CαOφ
3	CβOX	DαPφ	AδMΩ	BγNΨ
4	DγPΨ	CδOΩ	BαNφ	AβMX

Cars
A,B,C,D

Tires
M,N,O,P

Days
α,β,γ,δ

Speeds
φXΨΩ

The output (response) variable could be gas mileage for the 4 cars (A, B, C, D).

Fractional Factorial Experiments

The ASQ BOK refers to two-level fractional factorial designs. The authors have chosen to include one three-level example and a further explanation of EVOP in this topic area.

Two-Level Fractional Factorial Example*

The basic steps for a two-level fractional factorial design will be examined via the following example. The following seven step procedure will be followed:*

1. Select a process
2. Identify the output factors of concern
3. Identify the input factors and levels to be investigated
4. Select a design (from a catalogue, Taguchi, self created, etc.)
5. Conduct the experiment under the predetermined conditions
6. Collect the data (relative to the identified outputs)
7. Analyze the data and draw conclusions

CSSBB Test Success

Step 1: Select a process

Our ASQ Section wants to investigate CSSBB exam success using students of comparable educational levels.

Step 2: Identify the output factors

Student performance will be based on two results (output factors):

(1) Did they pass the test?
(2) What grade score did they receive?

Step 3: Establish the input factors and levels to be investigated

Our ASQ Section wants to study the effect of seven variables at two-level that they suspect may effect CSSBB student performance.

* This treatment of experimental design is the product of Mike Johnson, Plan Test Associates, 3012 North 32nd Street, Suite 22, Phoenix, AZ 85018. Mr. Johnson conducts seminars on SDE and can be contacted at 602-956-0180.

CSSBB Test Success (Continued)

Step 3: Inputs and levels (continued) (7 factors at 2-levels)

Input Factors	Level 1 (-)	Level 2 (+)
A. Refresher Course	No	Yes
B. Study Time	Morning	Afternoon
C. Problems Worked	200	800
D. Primary Reference	Book A	Book B
E. Method Of Study	Sequential	Random
F. Work Experience (Time)	4 Years	12 Years +
G. Duration Of Study	50 Hours	135 Hours

Note: The above inputs are both variable (quantitative) and attribute (qualitative).

Step 4: Select a design

A screening plan is selected from a design catalogue. Only eight (8) tests are needed to evaluate the main effects of all 7 factors at 2-levels. The design is:

	Input Factors						
Test	A	B	C	D	E	F	G
#1	-	-	-	-	-	-	-
#2	-	-	-	+	+	+	+
#3	-	+	+	-	-	+	+
#4	-	+	+	+	+	-	-
#5	+	-	+	-	+	-	+
#6	+	-	+	+	-	+	-
#7	+	+	-	-	+	+	-
#8	+	+	-	+	-	-	+

CSSBB Test Success (Continued)

Step 4: Select a design (continued)
One test example:

Factors	A	B	C	D	E	F	G
Test #3	-	+	+	-	-	+	+

Test #3 means:

A (-) = No refresher course
B (+) = Study in afternoon
C (+) = Work 800 problems
D (-) = Use reference book A
E (-) = Use sequential study method
F (+) = Have 12 years + of work experience
G (+) = Study 135 hours for the test

Step 5: Conduct the experiment

Step 6: Collect the data

Input Factors								Outputs (Test Results)	
Test	A	B	C	D	E	F	G	Pass	Score
#1	-	-	-	-	-	-	-	no (-)	31
#2	-	-	-	+	+	+	+	no (-)	59
#3	-	+	+	-	-	+	+	yes (+)	74
#4	-	+	+	+	+	-	-	no (-)	56
#5	+	-	+	-	+	-	+	yes (+)	87
#6	+	-	+	+	-	+	-	no (-)	69
#7	+	+	-	-	+	+	-	no (-)	44
#8	+	+	-	+	-	-	+	yes (+)	72

Avg 61.5

Step 7: Analyze the data and draw conclusions

The pass/fail pattern of (+)s and (-)s does not track with any single input factor. It visually appears that there is some correlation with factors C and G.

CSSBB Test Success (Continued)

What Factors Are Critical to Passing?

Input Factors							
	A	B	C	D	E	F	G
Test #1	-31	-31	-31	-31	-31	-31	-31
Test #2	-59	-59	-59	59	59	59	59
Test #3	-74	74	74	-74	-74	74	74
Test #4	-56	56	56	56	56	-56	-56
Test #5	87	-87	87	-87	87	-87	87
Test #6	69	-69	69	69	-69	69	-69
Test #7	44	44	-44	-44	44	44	-44
Test #8	72	72	-72	72	-72	-72	72

Difference (Δ)	52	0	80	20	0	0	92
Effect (Δ ÷ 4)	13	0	20	5	0	0	23

(+) means level 2 has a positive effect. (-) means level 2 has a negative effect. 0 means level 2 has no effect.

Step 7: Analyze the data and draw conclusions (continued)

- Factor A, taking refresher course, will improve the exam results by 13 points
- Factor B, study time of day, has no effect on exam results
- Factor C, problems worked, will improve the exam results by 20 points
- Factor D, primary reference, will improve the exam results by 5 points
- Factor E, method of study, has no effect on exam results
- Factor F, work experience, has no effect on exam results
- Factor G, duration of study, will improve the exam results by 23 points

CSSBB Test Success (Continued)

Step 7: Analyze the data and draw conclusions (continued)

To calculate the optimum student performance:

1. Sum the arithmetic value of the significant differences (Δ) and divide the total by two. Call this value the improvement. Note that the absolute value is divided by 2 because the experiment is conducted in the middle of the high and low levels and only one-half the difference (Δ) can be achieved.

A		B		C		D		E		F		G		
13	+	0	+	20	+	5	+	0	+	0	+	23	=	61

   Improvement = 61 ÷ 2 = 30.5. There were no significant negative effects (-) in this experiment. If there were, they would have been included (added) in determining the total effect. In this particular DOE format, the sign indicates direction only.

2. Average the test scores obtained in tests 1 through 8.

   Average = 61.5

3. Add the improvement to the average to predict the optimum performance.

   Optimum = Average + Improvement
   = 61.5 + 30.5
   = 92

   The optimum performance would be obtained by running the following trial:

A	B	C	D	E	F	G
+	0	+	+	0	0	+

   Where: + = Level 2   - = Level 1   0 = Doesn't matter

The above trial was one of the 120 tests not performed out of 128 possible choices. DOE is almost magical. Obviously, the predicted student scores can be confirmed by additional experimentation.

CSSBB Test Success (Continued)

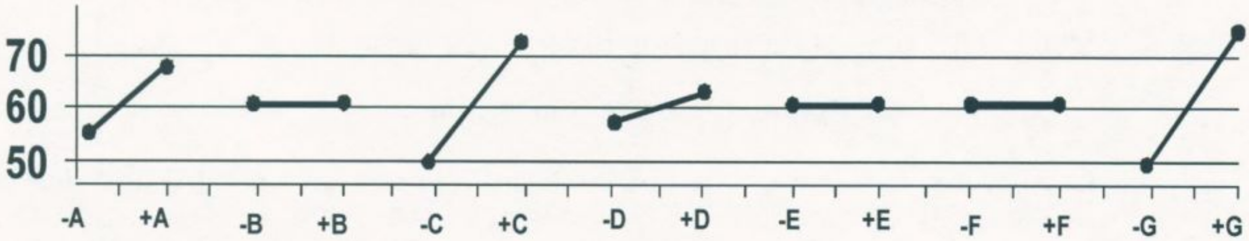
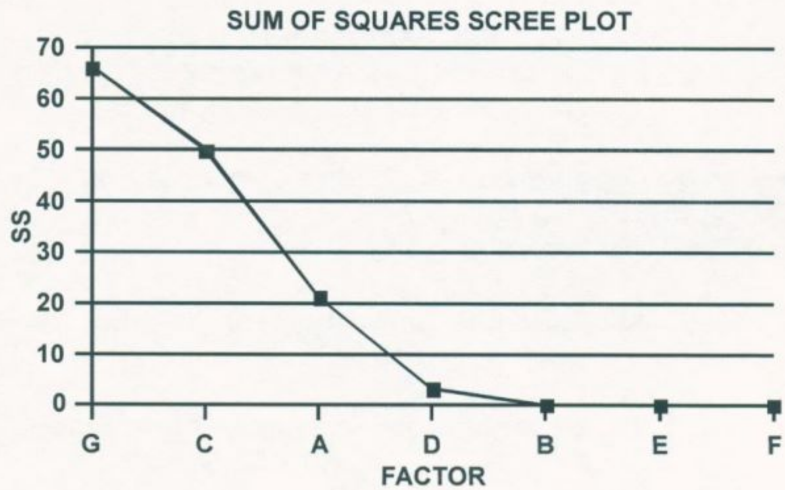


Figure 9.6 Plot of CSSBB Factor Levels vs. Test Scores

One can further examine the significance of the CSSBB design results using the sum of squares and a scree plot.

Note that $SS = \dfrac{(\Delta \text{ value})^2}{8}$

FACTOR	Δ	SS
G	23	66.1
C	20	50
A	13	21.2
D	5	3.1
B	0	0
E	0	0
F	0	0



A scree plot is so named because it looks like the rubble or rocky debris lying on a slope or at the base of a cliff. The scree plot indicates that factors D, B, E, and F are noise. The SS (sum of squares) for the error term is 3.1 (3.1 + 0 + 0 + 0).

$$\text{MSE (mean square error)} = \frac{3.1}{4} = 0.775$$

The maximum F ratio for factor G is: $\dfrac{66.1}{0.775} = 85.29$

The critical maximum F value from the following F Table for k - 1 = 7, p = 4 and α = 0.05 is 73. Thus, factor G is important at the 95% confidence level.

CSSBB Test Success (Continued)

The maximum F table accommodates screening designs for runs of 8, 12, 16, 20, and 24. p is the number of noise factors averaged to derive the MSE, and k is the number of runs.

The maximum F ratio for factor C is $\frac{50}{0.775} = 65.42$

The critical maximum F value for k - 1 = 7, p = 4 and α = 0.10 is 49. Thus, factor C is important at the 90% confidence level.

The maximum F ratio for factor A is $\frac{21.1}{0.775} = 27.22$

The critical maximum F values for both alpha values are larger than 27.22. Therefore, factor A is not considered important (at these alpha levels).

		8 Runs	12 Runs	16 Runs	20 Runs	24 Runs
p	alpha	k-1 = 7	k-1 = 11	k-1 = 15	k-1 = 19	k-1 = 23
2	0.05	974	3800	9000	17500	20200
	0.1	444	1590	3733	7950	9400
3	0.05	209	870	1838	4182	5198
	0.1	119	450	960	2273	2596
4	0.05	73	330	660	1618	2024
	0.1	49	194	424	1080	1236
5	0.05		155	355	812	975
	0.1		106	234	513	650
6	0.05		85	201	513	600
	0.1		57	135	332	408
7	0.05			126	308	391
	0.1			90	213	273
8	0.05			88	208	273
	0.1			62	150	192
9	0.05				160	195
	0.1				114	141

Table 9.7 Percentiles of the Maximum F Distribution for Screening Designs*
*** Adapted from (Wheeler, 1989)[23]**

Plackett-Burman Designs*

Plackett-Burman (1946)[15] designs are used for screening experiments. Plackett-Burman designs are very economical. The run number is a multiple of four rather than a power of 2.

Plackett-Burman geometric designs are two-level designs with 4, 8, 16, 32, 64, and 128 runs and work best as screening designs. Each interaction effect is confounded with exactly one main effect.

All other two-level Plackett-Burman designs (12, 20, 24, 28, etc.) are non-geometric designs. In these designs a two-factor interaction will be partially confounded with each of the other main effects in the study. Thus, the non-geometric designs are essentially "main effect designs," when there is reason to believe that any interactions are of little significance. For example, a Plackett-Burman design in 12 runs may be used to conduct an experiment containing up to 11 factors. See Table 9.8.

	Factors											
Exp	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	Results
1	+	+	+	+	+	+	+	+	+	+	+	
2	-	+	-	+	+	+	-	-	-	+	-	
3	-	-	+	-	+	+	+	-	-	-	+	
4	+	-	-	+	-	+	+	+	-	-	-	
5	-	+	-	-	+	-	+	+	+	-	-	
6	-	-	+	-	-	+	-	+	+	+	-	
7	-	-	-	+	-	-	+	-	+	+	+	
8	+	-	-	-	+	-	-	+	-	+	+	
9	+	+	-	-	-	+	-	-	+	-	+	
10	+	+	+	-	-	-	+	-	-	+	-	
11	-	+	+	+	-	-	-	+	-	-	+	
12	+	-	+	+	+	-	-	-	+	-	-	

Table 9.8 Plackett-Burman Non-geometric Design (12 Runs/11 Factors)

With a 20-run design, an experimenter can do a screening experiment for up to 19 factors. As many as 27 factors can be evaluated in a 28-run design.

* Source NIST (2001)[12]. Used with permission.

A Design from a Design Catalogue

The preferred DOE approach examines (screens) a large number of factors with highly fractional experiments. Interactions are then explored or additional levels examined once the suspected factors have been reduced.

A one-eighth fractional factorial design is shown below. A total of seven factors are examined at two-levels. In this design, the main effects are independent of interactions and six independent, two-factor interactions can be measured. This design is an effective screening experiment. This particular design comes from a design catalogue. Often experimenters will obtain a design generated by a statistical software program. Since this is a one-eighth fractional factorial, there are seven other designs that would work equally as well.

Test	Factors						
	A	B	C	D	E	F	G
1	0	0	0	0	0	0	0
2	0	0	0	1	1	1	0
3	0	0	1	0	1	1	1
4	0	0	1	1	0	0	1
5	0	1	0	0	0	1	1
6	0	1	0	1	1	0	1
7	0	1	1	0	1	0	0
8	0	1	1	1	0	1	0
9	1	0	0	0	1	0	1
10	1	0	0	1	0	1	1
11	1	0	1	0	0	1	0
12	1	0	1	1	1	0	0
13	1	1	0	0	1	1	0
14	1	1	0	1	0	0	0
15	1	1	1	0	0	0	1
16	1	1	1	1	1	1	1

Often, a full factorial or three-level fractional factorial trial (giving some interactions) is used in the follow-up experiment.

Note: 0 = low level and 1 = high level.

A Three-Factor, Three-Level Experiment

Often, a three-factor experiment is required after screening a large number of variables. These experiments may be full or fractional factorial. A one-third fractional factorial design is shown below. Generally the (-) and (+) levels in two-level designs are expressed as 0 and 1 in most design catalogues. Three-level designs are often represented as 0, 1, and 2.

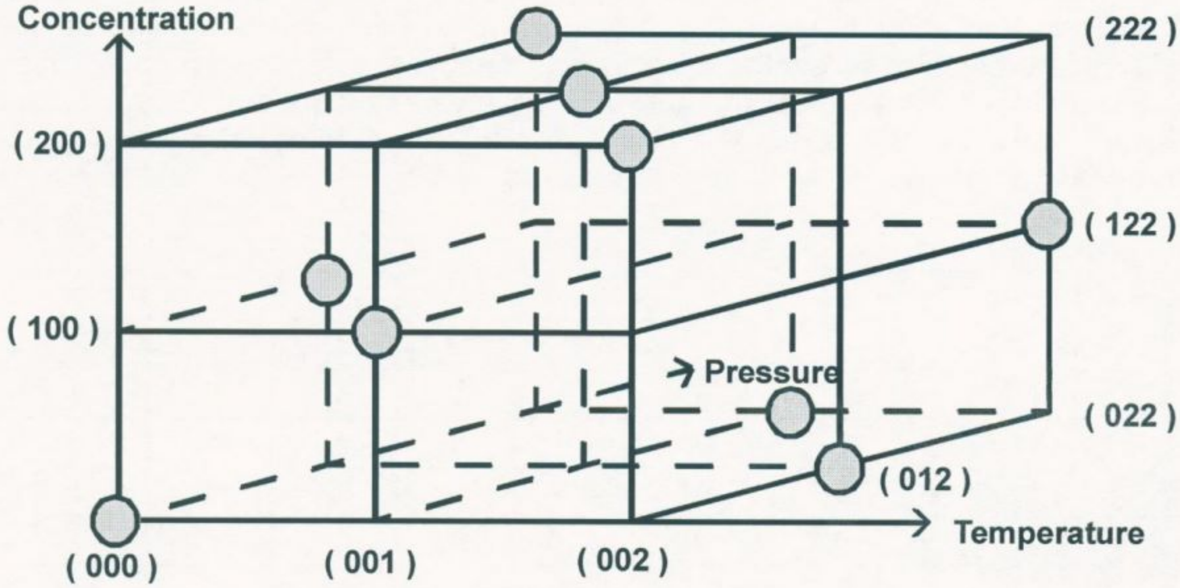


Figure 9.9 A one-third Fractional Factorial Design, 3-Factors, 3-Levels

From a design catalogue test plan, the selected fractional factorial experiment looks like so:

Experiment	Concentration	Pressure	Temperature
1	0	0	0
2	0	1	2
3	0	2	1
4	1	0	1
5	1	1	0
6	1	2	2
7	2	0	2
8	2	1	1
9	2	2	0

EVOP Evolutionary Operations

EVOP (evolutionary operations) emphasizes a conservative experimental strategy for continuous process improvement. Refer to Figure 9.10 below:

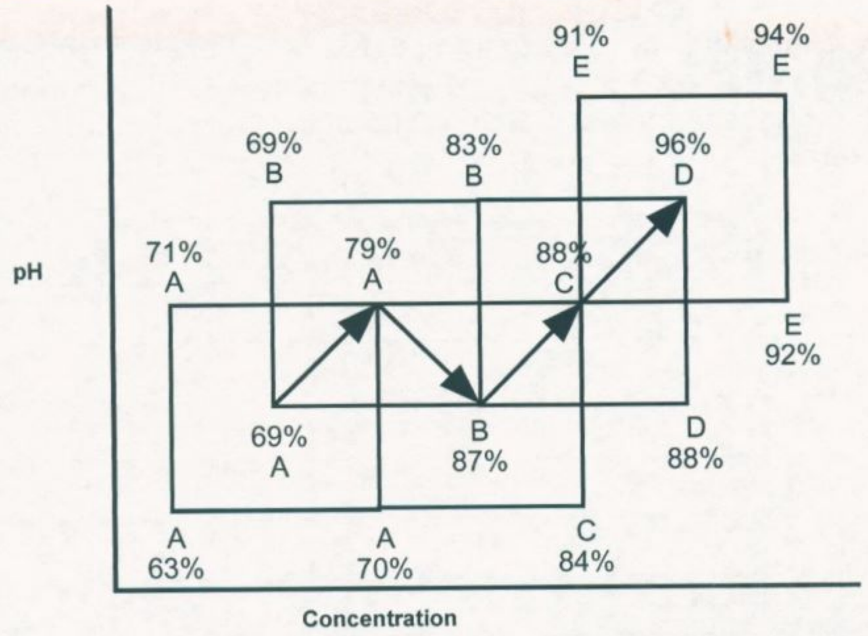


Figure 9.10 Illustration of EVOP Experimentation

Tests are carried out in phase A until a response pattern is established. Then phase B is centered on the best conditions from phase A. This procedure is repeated until the best result is determined. When nearing a peak, the experimenter will then switch to smaller step sizes or will examine different variables. EVOP can entail small incremental changes so that little or no process scrap is generated. Large sample sizes may be required to determine the appropriate direction of improvement. The method can be extended to more than two variables, using simple main effects experiment designs.

The experiment naturally tends to change variables in the direction of the expected improvement, and thus, follows an ascent path. In EVOP experimentation there are few considerations to be taken into account since only two or three variables are involved. The formal calculation of the direction of the steepest ascent is not particularly helpful.

A Full Factorial Example

Suppose that pressure, temperature, and concentration are three suspected key variables affecting the yield of a chemical process which is currently running at 64%. An experimenter may wish to fix these variables at two-levels (high and low) to see how they influence yield. In order to find out the effect of all three factors and their interactions, a total of $2 \times 2 \times 2 = 2^3 = 8$ experiments must be conducted. This is called a full factorial experiment. The low and high levels of input factors are noted below by (-) and (+).

Experiment No.	Temperature	Pressure	Concentration	% Yield
1	-	-	-	55
2	+	-	-	77
3	-	+	-	47
4	+	+	-	73
5	-	-	+	56
6	+	-	+	80
7	-	+	+	51
8	+	+	+	73
			Average	64

Temperature: (-) = 120°C (+) = 150°C
Pressure: (-) = 10 psi (+) = 14 psi
Concentration: (-) = 10N (+) = 12N

To find the effect of temperature, sum the yield values when the temperature is high and subtract the sum of yields when the temperature is low, dividing the results by four.

$$\text{The temperature effect:} = \frac{(77 + 73 + 80 + 73) - (55 + 47 + 56 + 51)}{4} = 23.5$$

A Full Factorial Example (Continued)

When the temperature is set at a high level rather than at a low level, one gains 23.5% yield. All of this yield improvement can be attributable to temperature alone since, during the four high temperature experiments, the other two variables were twice low and twice high.

$$\text{The pressure effect: } = \frac{(47 + 73 + 51 + 73) - (55 + 77 + 56 + 80)}{4} = -6$$

The effect of changing pressure from low level to high level is a loss of 6% yield.

$$\text{The concentration effect: } = \frac{(56 + 80 + 51 + 73) - (55 + 77 + 47 + 73)}{4} = 2$$

Higher concentration levels results in a relatively minor 2% improvement in yield.

The interaction effects between the factors can be checked by using the T, P, and C columns to generate the interaction columns by the multiplication of signs:

				Interactions				
EXP.	T	P	C	TXP	PXC	TXC	TXPXC	YIELD
1	-	-	-	+	+	+	-	55
2	+	-	-	-	+	-	+	77
3	-	+	-	-	-	+	+	47
4	+	+	-	+	-	-	-	73
5	-	-	+	+	-	-	+	56
6	+	-	+	-	-	+	-	80
7	-	+	+	-	+	-	-	51
8	+	+	+	+	+	+	+	73

Note, a formal analysis of the above data (developing a scree plot and MSE term) would indicate that only the temperature effect is significant. The scree plot and analysis was discussed earlier in this Section.

A Full Factorial Example (Continued)

Following the same principles used for the main effects:

$$\text{T x P interaction*: } = \frac{(55 + 73 + 56 + 73) - (77 + 47 + 80 + 51)}{4} = 0.5$$

***Interaction means the change in yield when the pressure and temperature values are both low or both high, as opposed to when one is high and the other is low.**

The T x P interaction shows a marginal gain in yield when the temperature and pressure are both at the same level.

$$\text{P x C interaction: } = \frac{(55 + 77 + 51 + 73) - (47 + 73 + 56 + 80)}{4} = 0$$

$$\text{T x C interaction: } = \frac{(55 + 47 + 80 + 73) - (77 + 73 + 56 + 51)}{4} = -0.5$$

$$\text{T x P x C interaction: } = \frac{(77 + 47 + 56 + 73) - (55 + 73 + 80 + 51)}{4} = -1.5$$

In this example, the interactions have either zero or minimal negative yield effects. If the interactions are significant compared to the main effects, they must be considered before choosing the final level combinations.

The best combination of factors here is: high temperature, low pressure, and high concentration (even though the true concentration contribution is probably minimal).

Comparison to a Fractional Factorial Design

In some situations an experimenter can derive the same conclusions by conducting fewer experiments. Suppose the experiments cost $10,000 each, one might then decide to conduct a one-half fractional factorial experiment.

A Full Factorial Example (Continued)

Comparison to a Fractional Factorial Design (Continued)

Assume the following balanced design is chosen.

Experiment	T	P	C	Yield
2	+	-	-	77
3	-	+	-	47
5	-	-	+	56
8	+	+	+	73

Since a fractional factorial experiment is being conducted, only the main effects of factors can be determined. Please note that experiments 1, 4, 6, and 7 would have been equally valid.

$$\text{The temperature effect} = \frac{(77 + 73) - (47 + 56)}{2} = 23.5$$

$$\text{The pressure effect} = \frac{(47 + 73) - (77 + 56)}{2} = -6.5$$

$$\text{The concentration effect} = \frac{(56 + 73) - (47 + 56)}{2} = 2.5$$

The results are not exactly identical to what was obtained by conducting eight experiments previously. But, the same relative conclusions as to the effects of temperature, pressure and concentration on the final yield can be drawn.

Note, the average yield is 63.25%. If the temperature is high, an 11.75% increase is expected, plus 3.25% for low pressure, plus 1.25% for high concentration equals an anticipated maximum yield of 79.5% even though this experiment was not conducted. This yield is in line with the actual results from experiment number 6 from the full factorial.

MINITAB Results

Most people don't analyze experimental results using manual techniques. The following is a synopsis of the effects of temperature, pressure, and concentration on yield results using MINITAB. This analysis represents the very same data for the previously presented examples.

Analysis of Variance for Yield versus T, P, C

Source	DF	SS	MS	F	p
T	1	1104.50	1104.50	803.27	0.000
P	1	72.00	72.00	52.36	0.002
C	1	8.00	8.00	5.82	0.073
Error	4	5.50	1.37		
Total	7	1190.00			

The F values and corresponding p-values indicate that temperature and pressure are significant to greater than 99% certainty. Concentration might also be important but more replications would be necessary to see if the 93% certainty could be improved to something greater than 95%.

Regression Analysis for Yield versus T, P, C

Predictor	Coef	SE Coef	t	p
Constant	-34.7500	6.4090	-5.42	0.006
T	0.7833	0.0276	28.34	0.000
P	-1.5000	0.2073	-7.24	0.002
C	1.0000	0.4146	2.41	0.073
S = 1.173	R^2 = 99.5%	R^2 (adj) = 99.2%		
The regression equation is: Yield = - 34.75 + 0.783 T - 1.50 P + 1.00 C				

The regression equation will yield results similar to those for the previous manual calculations. Again, the p-values for temperature and pressure reflect high degrees of certainty.

MINITAB Results (Continued)

Regression Analysis for Yield versus T, P, C
Fractional Factorial Example

Predictor	Coef	SE Coef	t	p
Constant	-36.7500	0.0000	*	*
T	0.7833	0.0000	*	*
P	-1.6250	0.0000	*	*
C	1.2500	0.0000	*	*
S = *	R^2 = *	R^2 (adj) = *		
The regression equation is: Yield = -36.75 + 0.783 T - 1.625 P + 1.250 C				

* indicates value cannot be calculated.

Using either the manual or MINITAB recaps, would the experimenter stop at this point? Might a follow-up experiment, perhaps at three levels looking at higher temperatures and lower pressures, pay off? After all, the yield has improved by 16% since experimentation started.

Experimental Notation

Often experiments are abbreviated numerically. For example, a 2^5 experiment means five factors at two levels (or 32 experiments). A mnemonic aid is: levels are level to the line, while factors fly.

Fractional designs can be expressed as L^{f-g}, where L is the number of levels, f is the number of factors, and g represents the number of generators. A generator of a fraction determines what effects are combined or confounded with one another. A design with g such generators is called a $1/L^g$ fractional factorial. Box (2005)[2]

For example a 2^{5-2} design is a two-level, five-factor, one-fourth fractional design. Rather than 32 runs, this experiment will require only eight runs.

Waste Elimination

Various lean manufacturing techniques are used in the six sigma control phase for waste elimination: 5S, visual controls, kaizen, kanban, poka-yoke, and standard work. Also important in this effort is the identification and minimization of the seven classic wastes reviewed in Section VIII. The application of many of these lean tools is summarized below.

5S (Housekeeping), Workplace Organization

Implementing 5S is a fundamental first step for any manufacturing company wishing to call itself world class. The presence of a 5S program is indicative of the commitment of senior management to workplace organization, lean manufacturing, and the elimination of muda (Japanese for waste). The 5S program mandates that resources be provided in the required location, and be available as needed to support work activities. The five Japanese "S" words for workplace organization are:

- Seiri (proper arrangement)
- Seiton (orderliness)
- Seiso (cleanup)
- Seiketsu (standardize)
- Shitsuke (personal discipline)

(Imai, 1997)[9], (Osada, 1991)[13]

For American companies, the translated English equivalents are:

- Sort: Separate out all that is unneeded and eliminate it
- Straighten: Put things in order, everything has a place
- Scrub (or shine): Clean everything, make the workplace spotless
- Standardize: Make cleaning and checking routine
- Sustain: Commit to the previous 4 steps and improve on them

The 5S approach exemplifies a determination to organize the workplace, keep it neat and clean, establish standardized conditions, and maintain the discipline that is needed to do the job. Numerous modifications have been made on the 5S structure. It can be reduced to 4S. It can be modified to a 5S + 1S or 6S program, where the sixth S is safety. The 5S concept requires that a discipline of will be installed and maintained.

5S (Continued)

Imai (1997)[9] relates the story of a Japanese team's initial site visit to a prospective supplier. Before allowing the supplier to unveil their grand presentation, the Japanese visitors insisted on a tour of gemba (the shop floor). After just a few minutes in the factory, the visitors knew that the plant was not committed to the highest level of manufacturing and terminated the visit. It is very easy to tell whether a plant is practicing a 5S program. In day-to-day operations, it is possible to have some dirt around the plant, but the visual signs of a 5S committed facility are obvious.

Details of a 5S program are itemized below in a step-by-step approach.

Step 1: Sort (Organize)

- Set up a schedule to target each area
- Remove unnecessary items in the workplace
- Red tag unneeded items, record everything that is thrown out
- Keep repaired items that will be needed
- Major housekeeping and cleaning is done by area
- Inspect the facility for problems, breakages, rust, scratches and grime
- List everything which needs repair
- Deal with causes of filth and grime
- Red tag grime areas and prioritize conditions for correction
- Perform management reviews of this and other steps

Step 2: Straighten

- Have a place for everything and everything in its place to ensure neatness
- Analyze the existing conditions for tooling, equipment, inventory and supplies
- Decide where things go, and create a name and location for everything
- Decide how things should be put away, including the exact locations
- Use labels, tool outlines, and color codes
- Obey the rules. Determine everyday controls and out-of-stock conditions
- Define who does the reordering and reduce inventories
- Determine who has missing items or if they are lost
- Use aisle markings, placement for dollies, forklift, boxes
- Establish pallet zones for work in process (WIP)

5S (Continued)

Step 3: Scrub (Shine and Clean)

- This is more than keeping things clean, it includes ways to keep things clean
- Establish a commitment to be responsible for all working conditions
- Clean everything in the workplace, including equipment
- Perform root cause analysis and remedy machinery and equipment problems
- Complete training on basics of equipment maintenance
- Divide each area into zones and assign individual responsibilities
- Rotate difficult or unpleasant jobs
- Implement 3-minute, 5-minute and 10-minute 5S activities
- Use inspection checklists and perform white glove inspections

Step 4: Standardize

- Make 5S activities routine so that abnormal conditions show up
- Determine the important points to manage and where to look
- Maintain and monitor facilities to ensure a state of cleanliness
- Make abnormal conditions obvious with visual controls
- Set standards, determine necessary tools, and identify abnormalities
- Determine inspection methods
- Determine short-term countermeasures and long-term remedies
- Use visual management tools such as color coding, markings and labels
- Provide equipment markings, maps, and charts

Step 5: Sustain

- Commit to the 4 previous steps and continually improve on them
- Acquire self-discipline through the habit of repeating the 4 previous steps
- Establish standards for each of the 5S steps
- Establish and perform evaluations of each step

(Imai, 1997)[9], (Osada, 1991)[13]

Management commitment will determine the control and self-discipline areas for an organization. A 5S program can be set up and operational within 5 to 6 months, but the effort to maintain world class conditions must be continuous. A well run 5S program will result in a factory that is in control.

Poka-Yoke / Mistake Proofing

Shigeo Shingo (1986)[19] is widely associated with a Japanese concept called poka-yoke (pronounced poker-yolk-eh) which means to mistake proof the process. Mr. Shingo recognized that human error does not necessarily create resulting defects. The success of poka-yoke is to provide some intervention device or procedure to catch the mistake before it is translated into nonconforming product.

Shingo (1986)[19] lists the following characteristics of poka-yoke devices:

- They permit 100% inspection
- They avoid sampling for monitoring and control
- They are inexpensive

Poka-yoke devices can be combined with other inspection systems to obtain near zero defect conditions. Errors can occur in many ways:

- Skipping an operation
- Positioning parts in the wrong direction
- Using wrong parts or materials
- Failing to properly tighten a bolt

(Suzaki, 1993)[22]

There are numerous adaptive approaches. Gadgets or devices can stop machines from working if a part or operation sequence has been missed by an operator. A specialized tray or dish can be used prior to assembly to ensure that all parts are present. In this case, the dish acts as a visual checklist. Other service-oriented checklists can be used to assist an attendant in case of interruption.

Numerous mechanical screening devices can be utilized. Applications can be based on length, width, height, and weight. Cash registers at many fast food outlets have descriptions or schematics of the product purchased. This system, in addition to the use of bar codes at supermarkets has eliminated data entry errors and saves time. Obviously, mistake proofing is a preventive technique.

Mistake proofing can also be accomplished through control methods by preventing human errors or by using a warning mechanism to indicate an error. Some of the control methods to prevent human errors include:

- Designing a part so it can not be used by mistake
- Using tools and fixtures that will not load a mis-positioned part
- Having a work procedure controlled by an electric relay

Poka-Yoke / Mistake Proofing (Continued)

A signaling mechanism can alert a worker of possible sources of error. Several applications include:

- Having the parts color-coded
- Having tool and fixture templates in place to only accept correct parts
- Having mechanisms to detect the insertion of a wrong part

A buzzer or light will signal that an error has occurred, requiring immediate action. Root cause analysis and corrective action are required before work resumes.

Other than eliminating the opportunity for errors, mistake proofing is relatively inexpensive to install and engages the operator in a contributing way. Work teams can often contribute by brainstorming potential ways to thwart error-prone activities. A disadvantage is, in many cases, that technical or engineering assistance is required during technique development.

Design improvements to mistake proof products and processes include:

- Elimination of error-prone components
- Amplification of human senses
- Redundancy in design (back-up systems)
- Simplification by using fewer components
- Consideration of functional and physical environmental factors
- Providing fail-safe cut-off mechanisms
- Enhancing product producibility and maintainability
- Selecting components and circuits that are proven

Everyday Examples of Poka-Yoke

- Gas cap attached to a car
- Gas pumps with automatic shut-off nozzles
- 110V electrical plugs and polarized sockets
- Microwave automatically stops when door is opened
- Seatbelt buzzer to warn drivers and passengers
- Elevator electric eye to prevent door from closing on people
- Lawn mower safety shut-off when bar is released
- Car keys ground symmetrical to allow two-way insertion
- Product drawings on cash registers at fast food restaurants
- Bar codes for product identification during distribution

Poka-Yoke / Mistake Proofing (Continued)

Figure 9.11 illustrates potential countermeasures for human errors.

Errors	Cause	Countermeasures
Forgetfulness	Poor concentration	Checklists/Visual aids
Misunderstanding	Unfamiliar situations	Training/Checklists Work standardization Visual aids/Work instructions
Identification	Similar appearance	Training/Visual aids
Amateur errors Beginner errors	Inexperience	Training/Skill building Work standardization Visual aids/Work instructions
Willful errors	Ignoring rules	Training/Work instructions
Inadvertent errors	Absent mindedness	Discipline/Skill building Work standardization Visual aids/Checklists
Slowness	Judgment delays	Work standardization Visual aids/Work instructions
Lack of standards	Inadequate instructions	Work standardization Work instructions
Surprise errors	Erratic equipment	TPM/Work standardization
Intentional errors	Crimes / Sabotage	Education/Discipline

Figure 9.11 Countermeasures for Human Errors

Poka-yoke techniques are especially effective when:

- Vigilance is required in manual operations
- Mis-positioning of components can occur
- Attributes not measurements are important
- SPC is difficult to apply
- Turnover and training costs are high
- Special cause failures occur frequently

Standard Work

The operation of a plant depends on the use of policies, procedures, and work instructions. These could be referred to as standards. Maintaining and improving standards leads to improvement of the processes and plant effectiveness.

If things go wrong in gemba, the workplace, such as defects or unhappy customers, efforts are made to seek out the root causes, implement corrective action, and change work procedures. If no problems are encountered in routine daily work (called maintenance), the process is under control. The first requirement of management is to maintain the standards. With a system under control, an improvement stage can be started. That is, management must not be satisfied with the status quo.

If there is a need for an increase in production, management must find a way to do so. One of the ways is to have operators change the way they do their jobs. The use of kaizen activities, kaizen blitz, etc., can be used to improve the process. Once the changes have been made, efforts should commence to standardize the new procedures. (Imai, 1997)[9]

Imai (1997)[9] provides a discussion of the term "standards." It seems that the word standards has a very bad connotation in the Western world versus that in Japan. In Japan, standards are used to control the process, not the workers. In the West, standards imply the use of unfair conditions on workers, such as working harder under extreme conditions, etc. In Japan, standards are used to describe a process that is the safest and easiest for the workers, and is the most cost-effective and productive way for the company. It is a balance between the two parties.

The following are some examples of standards that go beyond procedures and work instructions (Suzaki, 1993)[22]:

- Yellow lines on the floor
- Color coding
- Production control board
- Level indicators for minimum and maximum inventory
- Cross-training matrices
- Trouble lights

Standard Work (Continued)

Standard work is regarded to be one of the most important techniques for achieving a perfect process. This approach provides the discipline for attaining perfect flow in a process. Under normal work conditions, with no abnormalities in the system, the flow is perfect. The standard work conditions are determined for:

- Takt time
- Ergonomics
- Parts flow
- Maintenance procedures
- Routines

Standard work is the documentation of each action required to complete a specified task. Standard work should always be displayed at the workplace. (This is similar to the ISO 9001 standards which mention that work instructions should be accessible to the operator.) If abnormalities appear in the system, those items can be spotted and eliminated. (Sharma, 2001)[18]

According to Shingo (1989)[20], Toyota's Ohno mentioned that for standard operations, much of the information will be contained in standard work sheets. A standard work sheet combines the 3 elements of materials, workers, and machines in a work environment. Toyota refers to it as a work combination.

Standard work sheets, that operators will have confidence in, should consider the following:

- Resource availability
- Machine arrangements
- Process improvements
- Worker ideas valued
- Tooling improvements
- Minimized transport
- Optimized inventory
- Defective prevention
- Operational mistakes deterred
- Safe workplace concepts
- Autonomous systems installed

In order to have standard work sheets, waste elimination, problem solving, and quality methods must be accomplished. The elements that comprise the standard work operations are:

- Cycle time: The time allowed to make a piece of production. This will be based on the takt time. The actual time will be compared to the required takt time to see if improvements are needed.

Standard Work (Continued)

- Work sequence: The order of operations that the worker must use to produce a part: grasp, move, hold, remove, delay, etc. The same order of work must be done every time. A standard time is provided for each element. A standard work combination sheet (providing element times), standard work layout sheet (workplace layout), or planning capacity table (machine capacity data) or all 3 are provided to the operator for use.

- Standard inventory: This is the minimum allowable in-process material in the work area, including the amount of material on the machinery, needed to maintain a smooth flow. For continuous flow, one piece in the machine and one piece for hand off is optimal.

(Shingo, 1989)[20], (Sharma, 2001)[18]

Shingo (1989)[20] and Sharma (2001)[18] mentioned that standard charts will also include:

- Capacity charts by part: A chart on order of processes, process names, numbers, basic times, tooling needs, etc.

- Standard task combination sheets: Order of operations for a job.

- Task manuals: Detailed instructions on tool changes, setup changes, or parts assembly, etc.

- Task instruction manuals: A training manual for training workers. These have equipment layouts, quality check methods, operational procedures, standard stock required, etc.

- Standard operating sheets: Details from the task instruction manual with information on equipment layouts, cycle times, order of operations, standard on hand stock, net work times, safety checks, and quality checks.

Shingo (1989)[20] wrote that standard operating sheets should be improved continuously. Sharma (2001)[18] provides a monthly time length for review of work sheets. The proper implementation of standard work is not as exciting as performing "kaizen events." It is a time consuming process. The supervisors and workers must always be looking to update and maintain the standards. Sharma (2001)[18] provides a final definition of standard work:

> "The best combination of machines and people working together to produce a product or provide a service at a particular point in time."

Standard Work (Continued)

Standards have the following features:

1. Standards are the best, easiest, and safest way to do a job
2. They preserve employee know-how and expertise
3. They provide a way to measure performance
4. Correct standards show the relationship between causes and effects
5. Standards provide a basis for maintenance and improvement
6. They provide a set of visual signs on how to do the job
7. Standards are a basis for training
8. They are a basis for auditing
9. They are a means to prevent recurrence of errors
10. Standards minimize variability

(Imai, 1997)[9]

Takt Time

In the operation of a continuous flow manufacturing (CFM) line, takt time takes on great importance. Takt time is a term used (first by Toyota) to define a time element that equals the demand rate. In a CFM or one-piece flow line, the time allowed for each line operation is limited. The line is ideally balanced so that each operator can perform their work in the time allowed. The word taktstock is a German word for baton, used by an orchestra conductor (Imai, 1997)[9]. This provides a rhythm to the process, similar to a heartbeat. The German word takt means beat or stroke. The work pace is at a certain pace or rhythm. In CFM, a certain pace is maintained, and the line must be engineered to do so.

(Conner, 2001)[4], (Sharma, 2001)[18]

Kanban-Pull

T. Ohno of Toyota Motor Company was the originator of the kanban method. This idea supposedly occurred to Ohno on a visit to the United States when he visited a supermarket. In the supermarket, product is "pulled" from the shelf and the missing item is replenished. Liker (1997)[11] suggests that the story of Ohno visiting an American supermarket to develop Kanban is fiction.

Kanban is the Japanese word for "sign" and is a method of material control in the factory. It is intended to provide product to the customer with the shortest possible lead times. Inventory and lead times are reduced through Heijunka (leveling of production). For example, if the plant production goal for day 1 is 8 units of A and 16 units of B, and on day 2 it is 20 units of A and 10 units of B, the usual method is to produce all of A, followed by all of B. This may be the most efficient use of time for the plant machinery, but, since production will never go according to plan, the customer may change their mind on day 2 and order less of A. This causes a pileup of inventory and possibly increases cycle time.

To reduce the WIP and cycle time, the goal is to be able to produce each part, every day in some order such as, 2 As, 1 B, 2 As, 1 B, etc. The factory must be capable of producing such an arrangement. It requires control of the machinery and production schedule, plus coordination of the employees. (Liker, 1997)[11]

If a kanban system is used, with cards indicating the need to resupply, the method of feeding an assembly line could be achieved using the following process:

1. Parts are used on the assembly line and a withdrawal kanban is placed in a designated area.

2. A worker takes the withdrawal kanban to the previous operation to get additional parts. The WIP kanban is removed from the parts pallet and put in a specified spot. The original withdrawal kanban goes back to the assembly line.

3. The WIP kanban card is a work instruction to the WIP operator to produce more parts. This may require a kanban card to pull material from an even earlier operation.

4. The next operation will see that it has a kanban card and will have permission to produce more parts.

5. This sequence can continue further upstream.

Kanban-Pull (Continued)

The order to produce parts at any one station is dependent on receiving an instruction, the kanban card. Only upon receiving a kanban card will an operator produce more goods. This system aims at simplifying paperwork, minimizing WIP and finished goods inventories. Examples of kanban cards are shown below.

Production Instruction KANBAN		
Code		Color
RZC 5		
Type	Manual	
Quantity	1 Set	
Style	Standard	
Control	4M539ALR	

Production Instruction KANBAN		
Code		Color
MBT 8		
Type	Automatic	
Quantity	5 Each	
Style	Deluxe	
Control	1Z2673YQP	

Figure 9.12 Examples of Kanban Cards

Due to the critical timing and sequencing of a kanban system, improvements are continually made. A kanban system can not have production halted by machine failures or quality problems. Only a specific amount of product is in the system at any point in time. A stoppage will cause much distress throughout the production system. Every effort is made to eliminate causes of machine downtime and to eliminate sources of errors in production.

As the number of kanban cards is decreased, there is a resulting decrease in stock, and stoppages are again highlighted. All causes of stoppages must be eliminated, promoting production efficiency and improving quality. Shingo (1989)[20] notes that kanban systems are applicable in repetitive production plants, but not in one-of-a-kind production operations. Kanban is beneficial for production systems involving parts with common processes.

Kanbans are normally cards, but they can be flags or spaces on the floor, etc. All kanban details (such as part numbers, locations, bar code, delivery frequency, quantity, etc.) are forms of material control.

Cycle Time Reduction

Cycle time is a critical component of the operation of a lean enterprise. Cycle time is defined as the amount of time needed to complete a single task and to move it forward in the process. Cycle times may differ by task, but to make the line flow, all operations must be completed under a given takt time. A kaizen event or a line improvement activity can be utilized to analyze the operations and reduce the cycle times to be below the required takt time. The standard work is analyzed for value and non-value added work.

(Gee, 1996)[5], (Womack, 1996)[24]

A reduction in cycle time is undertaken for many of the following reasons:

- To please a customer
- To reduce internal or external wastes
- To increase capacity
- To simplify operations
- To reduce product damage (improve quality)
- To remain competitive

Training

Some of the cycle time training principles and topics are listed below:

- Introduction to the total systems concept
- Problem solving tools such as the "5 Whys"
- Importance of the next process as the customer
- Non-judgmental attitude to problem solving
- Identification of value and non-value added work
- Identification of muda (the seven wastes)
- Principles of motion study
- Work flow patterns (straight, T-shaped, U-shaped)
- Standard operations
- 5S workplace organization
- Visual management principles
- Just-in-time (JIT) production
- Poka-yoke principles
- Team dynamics

(Gee, 1996)[5]

Cycle Time Reduction (Continued)

To discuss the concept of cycle time reduction, consider an example of a line with 5 stations and observed cycle times as provided in Table 9.13.

Station	Work Time (seconds)
1	45
2	40
3	60
4	70
5	50
Total	265

Table 9.13 Work Time Example

The process required a takt time of 60 seconds, however only station 3 is at 60 seconds. It is good that stations 1, 2, and 5 are each below 60 seconds. Station 4 is above the takt time, so something must be done to reduce the cycle time there. An improvement team is observant enough to conclude that with a total cycle time of 265 seconds (or 66.25 seconds per station) a reduction in stations (or operators) might be achieved. One idea is to have 4 stations at 60 seconds each. The options are illustrated in Table 9.14.

Station	Work Time (seconds)	Option 1	Option 2	Option 2 after Kaizen
1	45	53	66.25	60
2	40	53	66.25	60
3	60	53	66.25	60
4	70	53	66.25	60
5	50	53		
Total	265	265	265	240

Table 9.14 Work Time Example Options

Cycle Time Reduction (Continued)

A kaizen team could be empowered to improve the 5 station line. The team facilitator would guide the team to a significant cycle time improvement within 5 days using a kaizen event. There is a general format to the kaizen event:

- There is a time limit of 5 days to accomplish the change (some use 3 days)
- Two days of training are provided on lean manufacturing techniques
- Two and one-half days are allotted for collecting data and making changes
- The last one-half day consists of a presentation on the results to the workforce (Gee, 1996)[5]

Team members actively participate in collecting and analyzing data. The operators, technical staff, supervisors, and maintenance staff can all be team members and be involved in the analysis. The team will perform work sampling, pace studies, line balancing, elemental analysis, motion studies, and takt time calculations.

Work sampling provides a picture of the work content of the station. This reveals the content and ratio of work, inspection, walking, and other factors. See Table 9.15 for content ratios.

Station	Work	Inspection	Delay	Walking	Other
1	75	5	10	10	
2	80	5	10	5	
3	65	5	5	20	5
4	50	5	5	35	5
5	65	15	5	10	5
Total	335	35	35	80	15

Table 9.15 Work Content Ratio (%)

The line value added activities, at this point, only comprise 50% to 80% of the work content of each station. The inspection, delay, walking, and other categories are considered muda. This provides the team with information to consider for station combinations. The team will investigate ways to eliminate (or reduce) the four muda elements. It appears possible to reduce the number of stations from five to four.

Cycle Time Reduction (Continued)

Further analysis of the work content ratio is displayed in Table 9.16.

Station	Original station time (seconds)	Work content (%)	Actual work time (seconds)
1	45	75	34
2	40	80	32
3	60	65	39
4	70	50	35
5	50	65	33
Total	265	65.3	173

Table 9.16 Main Work Analysis

As indicated in Table 9.16, the actual work time for the 5 stations amounts to 173 seconds. If there is presently 1 operator per station; given that the takt time is 60 seconds; the 173 seconds divided by 60 seconds suggests that 3 operators will be sufficient. There may still be a need for 5 stations, but only 3 operators. Additional data collection will be necessary to confirm this analysis. Pace studies of each station will provide a clearer picture of the cycle times. Usually up to 25 cycles of the line are studied in order to determine the average cycle time. A line balancing chart can be made and compared to the desired takt time. (Gee, 1996)[5]

A study of the stations reveals the motions used by the operators. In this study, an exacting industrial engineering approach to human motions is not used. An approximation of the operator effort will suffice. Robinson (1990) describes the Shingo technique of classifying human motions. It is divided into 4 grades:

1. Assemble, disassemble, and use (true value added)

2. Transport empty, grasp, transport loaded, and release load (non-value added)

3. Search, find, select, reposition, hold, inspect, and pre-position (non-value added, lower on the grading scale than above)

4. Rest, frequent planning, unavoidable delays, avoidable delays (non-valued added elements of the lowest rank)

Cycle Time Reduction (Continued)

The concept in studying human motions is to reduce the stress and strain upon the operators, creating a more efficient operation. The task activity of the operator can be classified and motion elements eliminated or reduced. Perhaps some mechanical device can be used to eliminate a non-valued element.

The team will prepare a workplace layout of the line. This layout will include operators, WIP inventory, raw materials, and equipment in the workspace. A charting of the current flow of the product may reveal a "spaghetti-like" flow. Thus, it can be termed a spaghetti chart. In many cases, rework and questions add many more lines than shown in Figure 9.17.

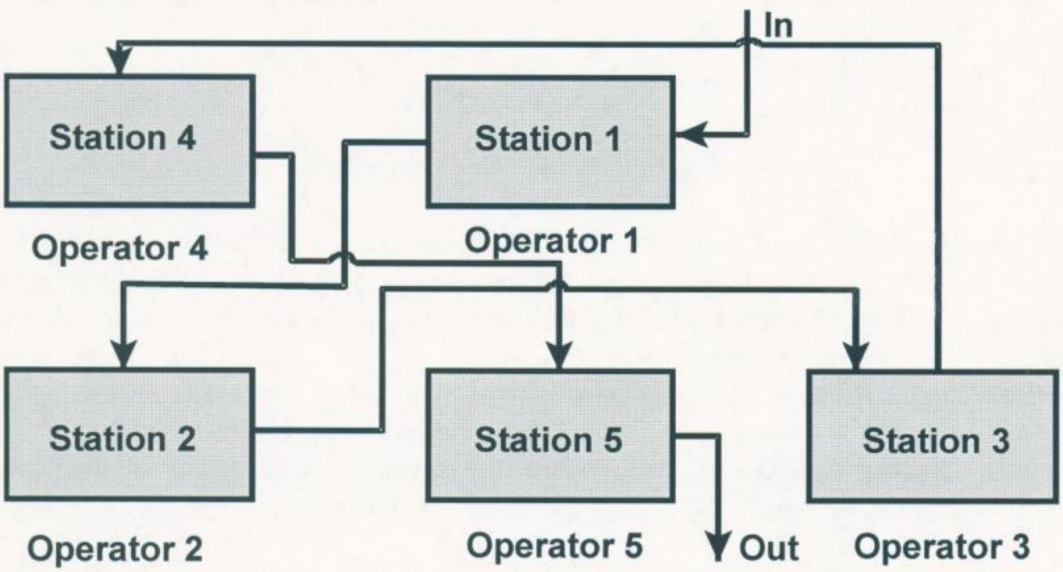


Figure 9.17 Flow Chart of Existing Production Line

The idea is to arrange the production line using either a U-shape, L-shape, C-shape, or straight line arrangement in order to create continuous flow. The various lines must reduce the distance traveled by the part, reduce the amount of WIP inventory between stations, and still meet the required takt time.

Cycle Time Reduction (Continued)

Refer to Figure 9.17 in which there is 1 operator per station. The process can be improved to reduce the distance traveled and to promote a smoother flow of work. Perhaps a U-shaped line can be constructed as shown in Figure 9.18.

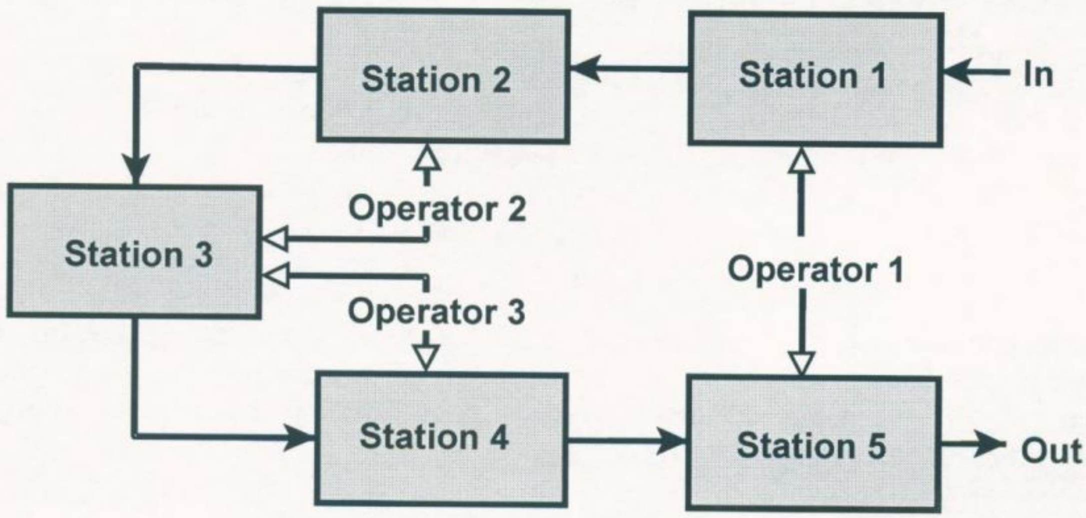


Figure 9.18 A Proposed U-Shaped Line

Figure 9.18 illustrates a U-shaped line with 3 operators. An analysis of the walking distance and material flow reveals significantly less wasted motion.

A kaizen event is a very stressful time for the whole team (especially for the facilitator), since no one at the beginning will know exactly how it will end up. There are many opportunities for innovation and creativity in the composition of line work load and layout. The final results are almost always pleasantly surprising in terms of achieving team goals. The team goals are usually:

- Reducing cycle times
- Meeting takt times
- Reducing space
- Reducing inventory
- Increasing line balance
- Maintaining a safe work environment

Continuous Flow Manufacturing (CFM)

In the lean thinking environment, continuous flow manufacturing is one of the basic principles. The main principle is that material should always be moving one-piece at a time, at a rate determined by the needs of the customer. The flow of product must be smooth and uninterrupted by:

- Quality issues
- Setups
- Machine reliability
- Breakdowns
- Distance
- Handling methods
- Transportation arrangements
- Staging areas
- WIP inventory problems, etc.

(Productivity, 1999)[16]

The mass production, or large lot production, world is a series of operations that produce goods in large batches. The sequence of operations used in producing parts in large batch sizes results in waiting time between operations. Large lot production has the following faults:

- Longer lead times for customer orders and delivery
- Additional resources in terms of labor, energy and space
- Additional product transportation expenses
- Increased product damage or deterioration costs

Continuous flow or one-piece flow will:

- Deliver a flow of products to the customer with less delay
- Require less storage and transport
- Lower the risk of losses through damage, deterioration, or obsolescence
- Provide a mechanism to solve other production problems

(Productivity, 1999)[16]

Continuous Flow Manufacturing (Continued)

Ideally, in a continuous flow manufacturing layout, the production steps are arranged in a tight sequence, such as a straight line or U-shaped cell, without WIP, using single piece flow. Inside this flow concept, each station and operator (in fact, the whole system) must operate with complete reliability to achieve continuous flow and the desired takt time. A system of production is attained with high quality levels using a variety of defect elimination and detection techniques.

A number of techniques, as discussed earlier in cycle time reduction are also useful for continuous flow manufacturing. The following concepts are important:

- Poka-yoke (mistake proofing): To prevent defects from proceeding to the next step.

- Source inspection: To catch errors that cause defects and to correct the process.

- Self-check: Checks by the operator to catch defects and to correct the process.

- Successive checks: Checks by the next process to catch errors and to correct the process.

- Total productive maintenance is used to help achieve high machine capability (see the discussion on TPM presented in Section X for more details).

(Womack, 1996)[24]

Setup Reduction

SUR is an acronym for setup reduction. SMED is an acronym for single minute exchange of dies. In this discussion, the two terms will be used interchangeably. SUR is one of the most important tools in the lean manufacturing system. The concept is to take a long setup change of perhaps 4 hours in length and reduce it to 3 minutes. Most people can not believe that this is possible. Shigeo Shingo, developer of the SMED system, used it quite effectively in the Toyota Production System for just-in-time production. Single minute exchange of dies does not literally require only one minute. It merely implies that die changes are to be accomplished in a single digit of time (nine minutes or less).

Long setup changes present a major problem for many companies dealing with low volume production. American industry has long held the view that the best production scheme would be long production runs of the same product. Witness Henry Ford's Model T assembly line of only black cars. In today's world, that may not be possible. There may be certain industries where the supplier is dominant in comparison to the buyer. However, if the customer is more dominant, or the industry is very competitive, being able to switch production very quickly can create a competitive edge. There are 3 myths regarding setup times:

- The skill for setup changes comes from considerable practice and experience
- Long production runs are more efficient because they save setup times
- Long production runs are economically better

SUR systems reduce the dependence on the long-term experience of operators to perform an effective changeover. Traditional methods depend on the setup operator having full knowledge of the machinery and components, and unique skills in mounting, removing, measuring, centering, adjusting, and calibrating setups. SUR systems reduce the skill level needed for setup changes. Long runs will reduce problems with setup changes, but lead to excess inventories, extra handling, extra storage, etc. The reduction of setup times:

- Expands production capacity
- Reduces inventories and minimize wastes
- Prevents quicker responses to demand changes
- Increases operating flexibility
- Makes more effective use of floor space
- Improves the utilization of capital equipment
- Reduces material handling
- Increases operator efficiency and safety (in some cases)

Setup Reduction (Continued)

To achieve quick changeover one determines operations that must be done while the machine is stopped, called internal setup (IS), and distinguish these from those which can be done while the machine runs, called external setup (ES). Any useless steps are removed. The sequence consists of:

- Removing useless operations
- Converting IS to ES
- Simplifying fittings and installations
- Suppressing adjustments and trials
- Working continually on improvement ideas

Quick changeover methods involve the following sequence of actions:

- Document all elements of current setup
- Separate internal from external operations
- Convert internal setups to external setups
- Generate ideas for reducing external setups
- Generate ideas for reducing internal setups
- Evaluate/test new ideas
- Prepare for the next changeover using new ideas
- Standardize by documenting actions and new procedures
- Continuously improve the process

The traditional setup method requires that all machine operations be stopped and then operators proceed to think about the setup. A better way is to identify what can be performed before shutting down the machine (external setup time), and then to identify what has to be done when the machine is shut down (internal setup time). A machine shut down means that no parts are being produced. In planning a SUR project, the actual conditions and steps of the die changeover must be detailed. This can be done by:

- Use of a stopwatch for continuous observation

- Use of a work sampling study

- Worker interviews

- Videotaping the entire setup operation

Setup Reduction (Continued)

After the first step has been performed, a breakdown of the operation can begin. The setup team may be comprised of operators, setup technicians, engineers, and maintenance staff. This team reviews the setup elements. Every step in the setup process (from start to finish) is broken down and classified. A major item is to separate items that can be done when the machine is running (external setup) and to separate the items that can only be done when the machine is down (internal setup). External setup operations should include:

- Preparation of parts
- Finding parts
- Measuring parts
- Maintenance of dies and spares
- Cleaning of spares, etc.

The break down of initial elements into internal and external setup operations is just a start. The existing internal setup elements should be reexamined to convert more of those elements into external setup. The goal is to reduce the time to under 1 digit. However, it may take a series of SUR projects to lower the time to 1 digit.

The setup team will need to generate some creative options. They should look for pre-heating of dies, earlier preparation of parts, simplifying holding devices, standardizing die heights, and using common centering jigs, multipurpose dies, parallel operations (2 or more people working), functional clamps, one-turn attachments, U-shaped washers, one-motion methods, interlocking methods, elimination of adjustments, etc. Brainstorming and problem solving sessions are needed to continuously improve the setup process.

All elements of internal and external setup must be reviewed in detail and streamlined in order to attain the single digit goal. Perhaps the goal is unattainable, but efforts should be made to go as low as possible. Once a SUR procedure is agreed upon, the setup team should practice the process and critique itself for additional improvements.

Setup Reduction (Continued)

Based on specific applications, the reduction of cycle time can have a considerable impact on the containment of costs and improvement in productivity. Consider the following case study, which is not a single minute exchange, but is representative of the process.

SUR Case Study

A casting facility in Virginia decided to investigate the time necessary to replace sand molds. In past years, the process had required 2 to 3 hours per change out. Through a gradual evolutionary process, that time had been reduced to an hour. A cross functional team was assigned the investigative task. Members included representatives of production, maintenance, engineering, and supervision. The team had previously undergone problem solving and team dynamics training.

Over a period of three months, the change out time was reduced to a firm fifteen minutes regardless of station and mold type. Some of the key ingredients in this success included:

- The mold storage method
- The staging of molds
- The timing of the change out
- A redesign of the die hardware

At least two minor engineering modifications were required. Plant communications, regarding the implementation of the new methods, were also part of the team's charter. Interestingly, the team discovered that their nearest competitor was still needing two hours to do the job. In this case, benchmarking the competitor (even if ethical) would have been discouraging.

Quick Response Manufacturing

According to Professor Ryjan Suri, quick response manufacturing (QRM) is the next step for the Toyota Production System (TPS). Ryjan Suri is a Professor of Industrial Engineering at the University of Wisconsin - Madison and Director of the Center for Quick Response Manufacturing. TPS is now 30 to 40 years old and could be considered an old technology. QRM helps companies use speed and the reduction of cycle times to deliver products and services faster than their competitors. This methodology can be applied to both the shop floor and the office.

In many cases, QRM requires that the managerial mind set must change. The implementation of QRM in a company requires proper training and orientation to grasp the dynamics of the manufacturing system. It is important to understand how capacity planning, resource utilization, lot-sizing, etc., interact with each other and impact lead times. This is very important in the relentless pursuit of lead time reductions. QRM is especially useful for a product line that has a large variety of highly engineered products with variable demand.

A specialized material planning technique that is a combination of both "push" and "pull" termed "POLCA", is used for controlling material flow. POLCA stands for Paired-cell Overlapping Loops of Cards with Authorization. This is a material control system that operates in conjunction with MRP and a cellular arrangement. Some examples of the benefits of QRM include:

- Reduction of lead times by 80% to 95%
- Lowered product costs by 15% to 30%
- Increased on-time deliveries from 60% to 99%
- Decreased scrap by 80%

(Suri, 2006)[21]

The QRM methodology focuses on speed. Suri highlights these principles:

- Change the management mindset
- Find ways to complete a job, focusing on lead time minimization
- Plan to operate critical resources at 70% to 80%, not 100%
- Use reduction of lead time not utilization as the main performance metric
- Do not use equipment efficiency or utilization as the main metrics
- Lead time reduction is more important than on-time delivery
- Install the POLCA material control system
- Move the suppliers to QRM
- Educate customers on QRM in order to enable smaller lot sizes
- Use quick response office cells (teams) for product families
- QRM will lead to a truly lean company

(Center for QRM, 2006)[3]

Kaizen

Kaizen is Japanese for continuous improvement (Imai, 1997)[9]. The word kaizen is taken from the Japanese kai "change" and zen "good." This is usually referred to as incremental improvement, but on a continuous basis, and involving everyone. Western management is enthralled with radical innovations. They enjoy seeing major breakthroughs, the home runs of business. Kaizen is an umbrella term for:

- Productivity
- Total quality control
- Zero defects
- Just-in-time
- Suggestion systems

(Imai, 1997)[9]

The kaizen strategy involves:

- Kaizen management: Management maintains and improves operating standards.
- Process versus results: Improvement of processes is the key to success.
- Use the PDCA/PDSA cycles: Plan-do-check-act is the method of improvement. The check cycle refers to verification that implementation has taken place and is on target to meet goals.
- Quality first: Quality is of the highest priority.
- Speak with data: Problems are solved with hard data.
- The next process is the customer: Every step of the process will have a customer. Provide the next step with good parts or information.

(Imai, 1997)[9]

Kaizen (Continued)

The Kaizen Blitz

While most kaizen activities are considered to be of a long-term nature by numerous individuals, a different type of kaizen strategy can occur. This has been termed a kaizen event, kaizen workshop, or kaizen blitz, which involves a kaizen activity in a specific area (involving planning, training, and implementation) within a short time period. (Gee, 1996)[5], (Laraia, 1999)[10]

The kaizen blitz, using cross functional volunteers in a 3 to 5 day period, results in a rapid workplace change on a project basis. The volunteers come from various groups, such as accounting, marketing, engineering, maintenance, quality and production. If the work involves a specific department, more team members are selected from that department.

Depending on the experience levels of the group, a 5 day kaizen blitz starts with 2 days of intense sessions on continuous improvement concepts. This is followed by 3 days of hands on data collection, analysis, and implementation at the source. The last portion of the workshop truly requires deep management commitment. Plant managers must trust the decision-making process as determined by the kaizen blitz team and facilitator.

A significant amount of time and money is involved at the implementation stage. The team makes a final presentation of the project to the plant manager and all interested plant employees. All project team members are encouraged to take part in the presentation. Every project has the possibility of bringing immediate changes and benefits. (Gee, 1996)[5]

Laraia (1999)[10] emphasizes that kaizen blitz events must occur with minimum expense and maximum use of people. The basic changes are in the process flow and methodology.

Various metrics are used to measure the outcomes of a kaizen blitz:

- Floor space saved
- Line flexibility
- Improved work flow
- Improvement ideas
- Increased quality levels
- Safe work environment
- Reduced non-value added time

Theory of Constraints

The theory of constraints (TOC) is a system developed by E. Goldratt. In 1986, Goldratt and Cox published a book titled *The Goal* (Goldratt, 1986)[7], which introduced the subject. *The Goal* describes a process of ongoing continuous improvement. Additional books have followed on the subject, including *Theory of Constraints* (Goldratt, 1990)[8].

Goldratt describes the theory of constraints as an intuitive framework for managing based on the desire to continually improve a company. Using TOC, a definition of the goals of the company are established along with metrics for critical measures.

(Goetsch, 2000)[6]

The Goal is a novel written in a story format describing the dual trials of a plant manager as he struggles to simultaneously manage his plant and his marriage. The key concept, "theory of constraints" is never mentioned as such, but is fed to the reader in bits and pieces. Listed below are many of the key elemental pieces:

- Bottlenecks
- Throughput
- Inventory
- Operational expenses
- Socratic way
- Jonah
- Common sense
- Delivery of results
- Goals
- Assumptions (most are incorrect)
- Return on investment
- Cash flow
- Local optimums
- Systems thinking
- Lead times
- Reduction of batch sizes
- Cost accounting
- Fear of change
- Resistance
- Net profit

The Goal reminds readers that there are three basic measures to be used in the evaluation of a system.

- Throughput
- Inventory
- Operational expenses

These measures are more reflective of the true system impact than machine efficiency, equipment utilization, downtime, or balanced plants.

Theory of Constraints (Continued)

A few of the most widely used TOC concepts are detailed below:

- Bottleneck resources are resources whose capacity is equal to or less than the demand placed upon it. A non-bottleneck is any resource whose capacity is greater than the demand placed on it. If a resource presents itself as a bottleneck, then things must be done to lighten the load. Some of the appropriate steps might be to offload material to relieve a bottleneck or to make the bottlenecks work only on parts needed now. One should beware of lost production at a bottleneck, due to poor quality or rejects.

- Balanced plants are not always a good thing. One should not balance capacity with demand, but balance the flow of product through the plant with demand from the market. The plant may be capable of generating inventories and goods at record levels, which jam up the plant's systems. The idea is to make the flow through the bottleneck equal to market demand. One can do more with less by just producing what the market requires at the time. It is possible that the existing plant has more than enough resources to do any job, but the flow must be controlled.

- Dependent events and statistical fluctuations are important. A subsequent event depends upon the ones prior to it. The story of Herbie and the local scout pack describes how the slowest member of a group will restrain the pace of the group. A bottleneck will restrain the entire throughput.

- Throughput is the rate at which the system generates money through sales. The finished product must be sold before it can generate money.

- Inventory is all the money that the system has invested in purchasing things that it intends to sell. This can also be defined as sold investments.

- Operational expenses are all the money that the system spends in order to turn inventory into throughput. This includes depreciation, lubricating oil, scrap, carrying costs, etc.

- The terms throughput, inventory, and operational expenses define money as incoming money, money stuck inside, and money going out.

(Goldratt, 1986)[7]

Theory of Constraints (Continued)

Goldratt (1990)[8] recommends that the following 5-step method be used for TOC implementation:

1. Identify the system's constraints. A system constraint limits the firm from achieving its optimum performance and goals. Thus, constraints must be identified and prioritized for maximum impact.

2. Decide how to exploit the system's constraints. The non-constraints in the system should be managed properly so that resources or materials are provided to feed the constraints.

3. Subordinate everything else to the above decisions. Constraints may have a limit, so look for ways to reduce the effects of a constraint, or look to expand the capacities of the constraints.

4. Elevate the system's constraints. Try to eliminate the problems of the constraint. Strive to keep improving the system.

5. Back to step 1. After the constraint has been broken, go back to step one and look for new constraints.

Drum - Buffer - Rope

The drum-buffer-rope concept relates to step 3 above: "Subordinate everything else to the above decisions." Most discussions on bottlenecks or constraints center on increasing the capacity or removing factors that slow the bottleneck. However, if the capacity of the bottleneck cannot be increased, then one must accept it, and then work to maximize the bottleneck's output.

One must ensure a smooth source of materials to the bottleneck. Because machines and equipment have variation in their output, there may, on occasion, be too much work-in-process (WIP) for the bottleneck, or not enough WIP for the bottleneck. The ideal situation is to always have enough WIP for the bottleneck (which controls the pace of the line) to keep production rates moving. Therefore, a set amount of inventory (a buffer) is needed ahead of the bottleneck.

Theory of Constraints (Continued)

Drum - Buffer - Rope (Continued)

To maintain a proper buffer level, a feedback mechanism is necessary to control the release of raw materials to the downstream equipment. One such technique is called the drum-buffer-rope concept, as described below:

- Drum: This is the constraint that controls the pace of the process. The "beat" of this operation sets the pace of the line.
- Buffer: This is the work-in-process, or inventory, for the bottleneck. It must be available to keep the bottleneck operating at full performance.
- Rope: This is the feedback mechanism from the buffer to the raw material input point. The dispatching point will release only enough material to keep the buffer inventory at the proper level.

The drum-buffer-rope technique can be thought of as a "pull-push" system. Refer to Figure 9.19 below.

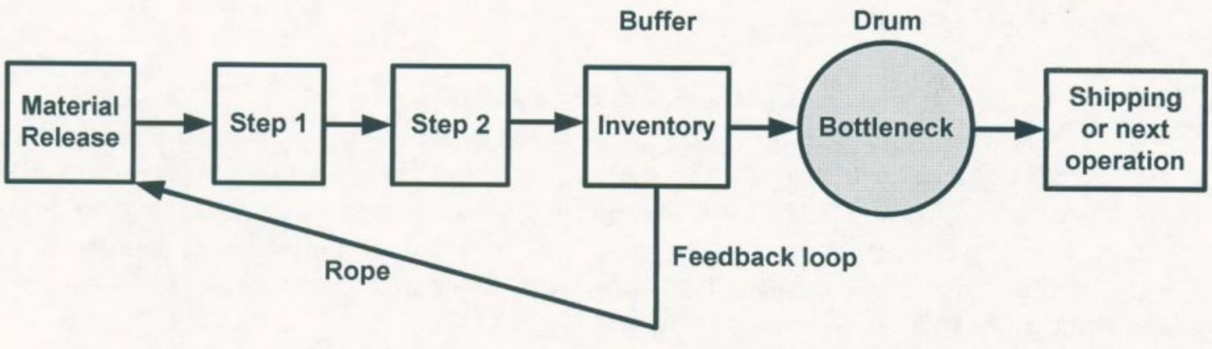


Figure 9.19 Drum - Buffer - Rope Example

Pitcher (2003)[14] indicates that the DBR model can work very well in a job shop with its wide variety of products, routings, and process times. In this environment, bottlenecks can be everywhere. The use of DBR methods has led to excellent performances in some situations, because WIP is kept low, and lower system cycle times are achieved.

(Pitcher, 2003)[14], (Yang, 2005)[26]

Theory of Constraints (Continued)

Goldratt (1990)[8] and Goetsch (2000)[6] describe the TOC approach as the Socratic method. This approach makes people find their own answers via the artful use of questioning. Direct answers are not given, but people are guided to draw their own conclusions and form their own opinions.

There are several other techniques used in the TOC environment:

- Effect-cause-effect: Use brainstorming to determine an intuitive sense of problems and their causes. That is, for an effect, provide assumptions for the effect, then speculate on a plausible cause. The cause is investigated to verify its validity.

- Evaporating clouds: Find the conflicting assumptions in the composition of a problem. Simple solutions are sometimes available for complex problems. Find the solution by reexamining the basic foundations of the problem. Once changes have been made to the system, the problem may no longer exist. It may evaporate.

- Prerequisite trees: Something must occur before something else can occur. TOC is a transition tool from an old way of doing things to a new way.

Management or improvement teams must determine the system constraint that is limiting production or throughput. A process chart is a good first step toward finding a constraint. An optional method is to develop a value stream map of the process which describes all the actions (both value added and non-value added) currently required to take a product from raw materials and deliver it to the customer. (Rother, 1999)[17]. Tree diagrams and process decision program charts are also highly effective tools.

Overall Equipment Effectiveness

Some discussion and calculations for OEE are reviewed in the TPM portion of Primer Section X.

Implementing Improved Processes

Listed below are a number of methods used to evaluate process and product improvements prior to full scale deployment.

Pilot run or pilot study: This is a trial of a changed product, process, equipment, or system to gain experience and collect data about the change. In the case of a proposed six sigma process change, operators (outside of the improvement team) are normally included both for training purposes and informational input.

Bothe (2002)[1] calls a pilot study a temporary introduction of a solution designed to confirm its effectiveness and uncover any potential problems with its eventual implementation. Normally this study is conducted for a limited time, in a limited area, by a limited number of people. Some common objectives of such a study include:

- Identifying potential implementation problems
- Discovering any adverse side effects
- Learning how to optimize the solution (Bothe, 2002)[1]

An improvement team should clearly state the objectives of the pilot study. In some cases, the study might uncover problems. In some cases, ways to reduce costs or increase the benefit of the proposed solution may result.

Simulation: This is a time-dependent trial of a new product, process, or system. This can often be accomplished by mathematical or computer-based modeling. This form of testing is conducted when an actual demonstration is too difficult, time consuming, expensive, massive, or dangerous. Normally parts, units, or operational people are not physically involved.

Demonstration: In the six sigma context, a demonstration can be an execution of an improved process or system. If a product is involved, there may be a requirement to obtain an approval for any real or potential changes. In the case of some products or processes, various human factors such as sight, touch, audio, thermal, or vibration may require investigation.

Prototype testing: This is the evaluation of a developmental model or unit that is close to production. It should be highly representative of the final equipment, parts, and processes. There may be additional manufacturing design changes but a prototype should allow for full mechanical and electrical evaluation.

References

1. Bothe, D. R. (2002). *Reducing Process Variation, Volume II.* Cedarburg, Wisconsin: Landmark Publishing Company.

2. Box, G. E., Hunter, W. G., Hunter, J. S., (2005) *Statistics for Experimenters,* 2nd ed. New York: John Wiley.

3. Center for Quick Response Manufacturing, University of Wisconsin - Madison. (2006). Retrieved November 22, 2006 from: http://www.engr.wisc.edu/centers/cqrm/pub/qrm21st.htm.

4. Conner, G. (2001). *Lean Manufacturing for the Small Shop.* Dearborn, MI: Society of Manufacturing Engineers.

5. Gee, G., McGrath, P., & Izadi, M. (Fall, 1996). "A Team Approach to Kaizen." Journal of Industrial Technology, 12 (4).

6. Goetsch, D. L. & Davis, S. B. (2000). *Quality Management, Introduction to Total Quality Management for Production, Processing, and Services,* 3rd ed. Upper Saddle River, NJ: Prentice Hall.

7. Goldratt, E., & Cox, J. (1986). *The Goal: A Process of Ongoing Improvement,* Revised Edition. Croton-on-Hudson, NY: North River Press.

8. Goldratt, E. (1990). *Theory of Constraints.* Great Barrington, MA: North River Press.

9. Imai, M. (1997). Gemba Kaizen: *A Commonsense, Low-Cost Approach to Management.* New York: McGraw-Hill.

10. Laraia, A., Moody, P., & Hall, R. (1999). *The Kaizen Blitz: Accelerating Breakthroughs in Productivity and Performance.* New York: John Wiley & Sons.

11. Liker, J. (editor). (1997). *Becoming Lean: Inside Stories of U.S. Manufacturers.* Portland, OR: Productivity Press.

12. NIST, (2001). *Engineering Statistics Handbook.* Retrieved from http://www.itl.nist.gov/div898/handbook/ September 21, 2001.

13. Osada, T. (1991). *The 5S's: Five keys to a total quality environment.* Minato-ku, Tokyo: Asian Productivity Organization.

References (Continued)

14. Pitcher, M. (March 13, 2003). "Examining Just-in-Time and Theory of Constraints." Retrieved November 29, 2006 from: http://www.sme.org/cgi-bin/get-newsletter.pl?lean&2003013&1.

15. Plackett, R. L. & Burman, J. P. (1946). "The Design of Optimal Multifactorial Experiments." Biometrika (Vol.33).

16. Productivity Development Team. (1999). *Cellular Manufacturing: One-Piece Flow for Workteams*. Portland, OR: Productivity Press.

17. Rother, M., & Shook, J. (1999). *Learning to See: Value Stream Mapping to Add Value and Eliminate Muda*. Brookline, MA: The Lean Enterprise Institute.

18. Sharma, A., & Moody, P. (2001). *The Perfect Engine: How to Win in the New Demand Economy by Building to Order with Fewer Resources*. New York: The Free Press.

19. Shingo, S. (1986). *Zero Quality Control: Source Inspection and the Poka-yoke System*. Stamford, CT: Productivity Press.

20. Shingo, S. (1989). *A Study of the Toyota Production System From an Industrial Engineering Viewpoint*. Cambridge, MA: Productivity Press.

21. Suri, R. (2006). Quick Response Manufacturing: A Competitive Strategy for the 21st Century. Retrieved November 23, 2006 from: http://www.advancedmanufacturing.com/may01/qrm.pdf.

22. Suzaki, K. (1993). *The New Shop Floor Management: Empowering People for Continuous Improvement*. New York: The Free Press.

23. Wheeler, D. J. (1989). *Understanding Industrial Experimentation*, 2nd ed. Knoxville, TN: SPC Press, Inc.

24. Womack, J., & Jones, D. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. New York: Simon & Schuster.

25. Wortman, B. L., et al. (2012). *CQE Primer*. Terre Haute, IN: Quality Council of Indiana.

26. Yang, K. (2005). *Design for Six Sigma for Service*. New York: McGraw-Hill.

