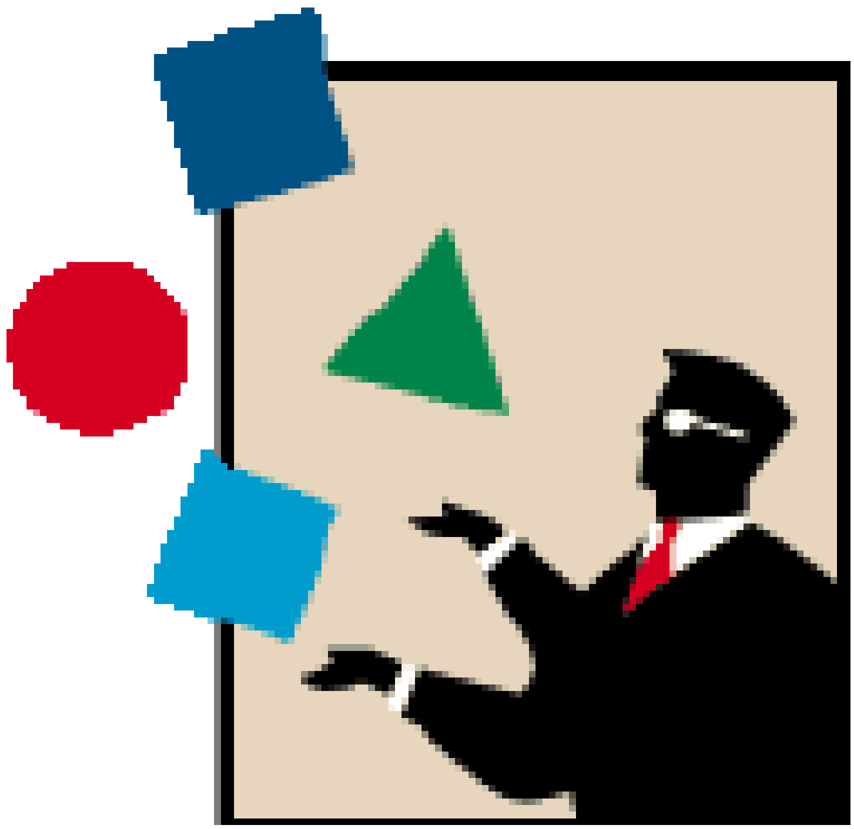
Strategy for Simultaneous
Evaluation of Multiple Objectives



Nutek, Inc. (Since 1987)

Quality Engineering Seminar and Software

Bloomfield Hills, MI, USA. 2014

Course Content

1. Introduction
2. Situations and needs for simultaneous evaluations
3. Technical difficulties and challenges in engineering practices
4. Single index formulation – an overall evaluation criterion
5. Example calculations
6. Application areas
7. References

1. Introduction

Strategy for Simultaneous Evaluation of Multiple Objectives

When a product or process is to satisfy multiple objectives that are evaluated by different criteria of evaluations, it is possible to combine them into a single index for analysis of overall performance. Intended for experienced experimenters interested in designing experiments to optimize multiple objectives simultaneously, this session will show you a scheme of handling evaluations of multiple objectives together by combining it into a single number called the Overall Evaluation Criteria (OEC). The OEC formulation can be used to combine and calculate the single index before analyzing the results. The OEC can be hand-calculated or computed using software capabilities before analyzing the results. You should consider attending this session, if your products or processes have multiple objectives and you are looking toward planning experiments to evaluate the results and determine design specifications based on the overall performance.

Objectives

Objectives are things you are after.
How do you measure it?
What is the quality characteristic (QC)?

Often we are after more than one single thing. When our objectives are different, it is likely that they all are measured with different units of measurement, and also they may have different sense of desirability.

2. Situations and needs for simultaneous evaluations

The need for simultaneous evaluation arises when there is more than one objective that a product or process is expected to satisfy. Situations of this nature are more common in many areas of our lives. Consider the educational system of rating students. Students are evaluated separately in each of the courses they take. But for comparison purposes, it is the Grade Point Average (GPA) which is used most often. Do you ever wonder where the performance numbers like 5.89, 4.92, etc. come from for the Olympic Figure Skating competition? Obviously these numbers are averages of all the judges' scores on a scale of 0 to 6. Each of the judges, of course, use the same range of evaluation numbers to evaluate different aspects of the performance like *style of skating, how high they jump, how well they land,* etc. The scores one judge assigns to a performer come from averaging his/her scores in all separate criteria of performance. Such use of an average or overall criteria of evaluation is quite common in many

Why do we need to look at them together?

- Evaluate overall performance
- Decision based on one objective may not satisfy the other objectives well
- Set up an impartial judge (system) to objectively decide performances

activities. It is not so common in engineering and science however.

Who does it?

- Educational system/organization uses grade point average (GPA) to evaluates students overall academic performance.
- Athletic events like figure skating in Olympic uses a single index (in a scale of 0 - 6) to evaluate performance of the competitors.

To evaluate several objectives together requires combining evaluations in several criteria into a single index, like GPA in case of students' achievement. An index like **Overall Evaluation Criteria (OEC)** can be formulated to evaluate performances of products and processes in engineering terms. The OEC allows evaluation of multiple objectives using a single numerical index which is formed by combining the different criteria of evaluations. But, it is not commonly done.

3. Technical difficulties and challenges in engineering practices

If an overall index is so common in sporting, educational, and social events, why is it so rare in engineering and scientific studies? It so primarily, because of three difficulties:

1. **Units of Measurements** - Unlike GPA or Figure Skating, the criteria of evaluations in engineering and science are generally different (say psi for pressure and inch for length, etc.). When units are different, they cannot be combined easily.

   Solution: Convert reading into a fraction by dividing by a *"fixed number*".

2. **Relative Weighting** - In formulating the GPA, all courses for the student is weighted the same. This is generally not the case in scientific studies. Consider, for example the case of baking POUND CAKES. There might be three objectives such as TASTE, MOISTNESS, and SMOOTHNESS that are important. But these criteria of evaluation may not all be of equal importance. How do you decide which one is important? This is generally done subjectively, by project team consensus.

   Solution: Multiply each reading (fraction) by the appropriate weighting

3. **Sense of the Quality Characteristic (QC)** - Quality characteristic (QC) indicates the direction of desirability of the evaluation numbers. Depending on the criteria and how it is measured, QC can be BIGGER IS BETTER, SMALLER IS BETTER, or NOMINAL IS THE BEST. For example in the game of Golf, a SMALLER score is better. In Basketball, on the other hand, a larger score is BETTER. Again, unlike the sporting and educational events, the difference of QC in scientific studies is quite common. Unless the quality characteristics of different criteria are the same, the evaluation numbers cannot be readily combined.

   Solution: Subtract reading fraction from "1" or a fixed number, to change the QC

Example of Criteria with Different QC: Determine the best player

	Player 1	Player 2	QC	
Golf (9 holes)	42	52	SMALLER	(average score)
Basketball	28	18	BIGGER	(individual score)
Total Score =	70	70		

So, why not do it in engineering?

There are three hurdles to overcome.

- Units of measurements
- Relative weightings of different criteria
- Sense of desirability (QC) of different criteria

Are these two players of equal caliber? Is the addition of the scores meaningful and logical? Obviously the total of scores has no meaning, as these players to do not perform equally. To make the total score meaningful, one of the numbers must be adjusted such that all QC's are aligned one way (either BIGGER or SMALLER).

A logical and meaningful way to combine the two scores will to first change the QC of the Golf score by subtracting from a fixed number (1 of the reading is in fraction), say 100 (an arbitrary number, expected highest possible strokes for 9 holes of Golf), then add it to the score of the Basketball. The new total score then becomes:

Total score for Player 1 = 30 + (100- 45) = 85

Total score for Player 2 = 20 + (100- 55) = 65

The number 85 and 65 now indicate the relative merit of the players, Player 1 (85) Player 2 (65).

Multiple objectives are quite frequent in engineering applications. No matter the project, be it a product optimization, process study, or problem solving, the desire to satisfy more than just one objective is quite common. Because the criteria involved are different, the experimental results (be it DOE/Taguchi or otherwise) are generally analyzed one criteria at a time. This approach, of course, does not guarantee that the best design obtained for one criterion, will also be desirable for the other criterion. What is needed is a properly formulated Overall Evaluation Criteria (OEC) number representing the performance of the test sample. Thus, when there are multiple criteria of evaluations, lack of such formulation poses a major hurdle for analysis of DOE results.

4. Single index formulation – an overall evaluation criterion

OEC Formulation

When the product or process under study is to satisfy more than one objective, performances of samples tested for each trial condition are evaluated by multiple criteria of evaluation. Such evaluations can be combined into a single quantity, called Overall Evaluation Criteria (OEC), which is considered as the result for the sample. Each individual criterion may have different units of measurements, quality characteristic, and relative weight. In order to combine these different criteria, they must first be normalized and weighted accordingly as shown below.

The method of OEC formulation and computation can be studied by considering the cake baking.

EVALUATION CRITERIA DESCRIPTION

Criteria	Worst Value(w)	Best Value(b)	QC	Weighting(Rw)
1.Taste	0	12	Bigger	55%(Rw1)
2. Moistness	25 gm	40 gm	Nominal	20% (Rw2)
3. Consistency	8	2	Smaller	25% (Rw3)

Required Qualities of a SINGLE INDEX

- Normalized – no units/fraction

  Fraction = (Reading / Reference value)

- QC's aligned

  Aligned Fraction =(1 - Fraction)

- Appropriately weighted

  Aligned Fraction x Weighting

Overall Evaluation Criteria (**OEC**) is that single index.

The evaluation criterion were defined such that TASTE was measured on a scale of 0 to 12 (bigger is better), MOISTNESS was indicated by weight with 40 gm (target value) considered the best value (nominal is better) and CONSISTENCY was measure in terms of the number of voids seen (smaller is better).

Assume that the cake sample for trial#1, the readings are (T, M, C):

Taste T = 9, Moistness M = 34.9, and Consistency C = 5

Then OEC for the cake sample is can be expressed as:

OEC = [l(9-0)l/(12-0)] x Rw1 + [1 - l(34.9 - 40)l/l(40-25)] x Rw2 + [1 - l(5-2)l/(8-2)] x Rw3

Explanation of data reduction:

Numerator *(9 - 0) represents (reading - worst value) in case of BIGGER QC*

Numerator *(40 - 34.9) represents (reading - target value) in case of NOMINAL QC*

Numerator *(5 - 2) represents (reading - best value) in case of SMALLER QC*

Denominators *(12 - 0, 40 - 25, and 8 - 2) represent differences between the best and the worst values for all QC. The worst value in case of NOMINAL is the worst deviant of the data extremes from the target.*

Note: Before all criteria of evaluations can be combined, their QC's must all be the same. The second expression is modified to change the NOMINAL QC, first SMALLER by finding the deviation from the nominal, then to BIGGER. The third expression is modified to change the SMALLER QC to BIGGER. The numerator in each term is calculated by subtracting the smaller magnitude (or target value in case of NOMINAL) from the reading, then taking the absolute (|x|) value. The denominator is always the range of data spread, which is positive difference between the best and the worst reading for the criteria.

$$OEC = \left[\frac{9}{(12-0)} \right] \times 55 + \left[1 - \frac{(40 - 34.9)}{(40-25)} \right] \times 20 + \left[1 - \frac{(5-2)}{(8-2)} \right] \times 25$$

$$OEC = \left[\frac{9}{12} \right] \times 55 + \left[1 - \frac{5.81}{15.0} \right] \times 20 + \left[1 - \frac{3}{6} \right] \times 25$$

$$= 41.25 + 12.25 + 12.50 = 66.00$$

Observe that the evaluation (T, M and C) in each case is first modified to show the positive difference between the reading and the smaller magnitude of the best/worst values (or target for NOMINAL), then divided by the allowable spread of the data. This is done to get rid of the associated units (Normalization). Subtraction of the fractional reading from 1, as done for the second and the third criteria, is to change the quality characteristics to line up with the first criteria, that is, BIGGER IS BETTER. Each criteria's fractional value (y/ymax) is also multiplied by the corresponding weighting and added together to produce a net result in numerical terms. The manipulation above normalizes OEC formulation to produce numbers between 0 - 100.

Since an L-8 orthogonal array described the baking experiment, there are eight cakes baked with one sample per trial condition.. The OEC calculated above (OEC = 66) represents the result for trial#1 . There will be seven other results like this. The eight values (OEC's) will then form the result column in the orthogonal array. The process will have to be repeated if there were more repetitions in each trial condition.

Example Formulation of OEC

Criteria	Worst Value(w)	Best Value(b)	QC	Weighting(Rw)	Sample Reading
1.Taste	0	12	Bigger	**55%(Rw1)**	**9**
2. Moistness	25 gm	40 gm	Nominal	20% (Rw2)	34.9
3. Consistency	8	2	Smaller	25% (Rw3)	5

Approach: Make OEC Bigger by changing all QC's to bigger

For Taste
Fraction = [(9 - 0) /(12 – 0)], No need to align, Weighted Fraction = (9/12) x 55

$$\textbf{Contribution} = \frac{(9 - 0)}{(12 - 0)} \times 55$$

Reading

Rel. Weighting

Reference value

Example Formulation of OEC

Criteria	Worst Value(w)	Best Value(b)	QC	Weighting(Rw)	Sample Reading
1.Taste	0	12	Bigger	**55%(Rw1)**	**9**
2. Moistness	25 gm	40 gm	Nominal	20% (Rw2)	34.9
3. Consistency	8	2	Smaller	25% (Rw3)	5

Approach: Make OEC Bigger by changing all QC's to bigger

For Moistness
Aligned Fraction = [1 – (40-34.9)/(40-25)], Weighted Fraction = [1 – (40-34.9)/(40-25)] x 20

Reading — Rel. Weighting

$$\textbf{Contribution} = \left[1 - \frac{(40-34.9)}{(40 - 25)} \right] \times 20$$

Target Value — Reference value

Note: (Nominal case)

1. The reading is reduced to difference from the target, which changes the QC to *smaller is better*.
2. The deviation reading and the reference number, maximum deviation, are always calculated positive (absolute values)
3. Subtract the aligned fraction from 1 to change QC.

Example Formulation of OEC

Criteria	Worst Value(w)	Best Value(b)	QC	Weighting(Rw)	Sample Reading
1.Taste	0	12	Bigger	**55%(Rw1)**	**9**
2. Moistness	25 gm	40 gm	Nominal	20% (Rw2)	34.9
3. Consistency	8	2	Smaller	25% (Rw3)	5

Approach: Make OEC Bigger by changing all QC's to bigger

For Moistness

Aligned Fraction = [1 – (40-34.9)/(40-25)], Weighted Fraction = [1 – (40-34.9)/(40-25)] x 20

$$\textbf{Contribution} = \left[1 - \frac{(5 - 2)}{(8 - 2)} \right] \times 25$$

Reading

Rel. Weighting

Reference value

[One – Fraction] to align QC

Example Formulation of OEC

Criteria	Worst Value(w)	Best Value(b)	QC	Weighting(Rw)	Sample Reading
1.Taste	0	12	Bigger	55%(Rw1)	9
2. Moistness	25 gm	40 gm	Nominal	20% (Rw2)	34.9
3. Consistency	8	2	Smaller	25% (Rw3)	5

Approach: Make OEC Bigger by changing all QC's to bigger

For Taste
Fraction = [(9 - 0) /(12 – 0)], No need to align, Weighted Fraction = (9/12) x 55

For Moistness
Aligned Fraction
= [1 – (40-34.9)/(40-25)], Weighted Fraction = [1 – (40-34.9)/(40-25)] x 20

For Consistency
Aligned Fraction
= [1 – (5-2)/(98-2)], Weighted Fraction = [1 – 5/(98-2)] x 25

OEC = (9/12)x55 + [1 – (40-34.9)/(40-25)]x20 + [1 – 5/(98-2)]x25

= 41.25 + 12.25 + 12.50

= 66.00

5. Example calculations

Notations for OEC Formula

X = Reading/evaluation of sample (for the criterion concerned)
X_L = Largest value of reading,
X_S = Smallest value of reading,
X_N = Nominal/Target value for criteria with nominal QC.
B = Bigger is better
S = Smaller is better
QC = Quality characteristics

Steps for OEC Formulation

a. Prepare criteria description table
b. Determine relative weighting by group consensus when possible (or assume)
c. Select QC for the OEC based on relative weights of the criteria (Select BIGGER, unless all criteria weighting with smaller QC exceeds 60%)
d. Based on above (item c.), write OEC equation with proper adjustment for change of criteria QC
e. Transform readings by subtracting smaller magnitude of the range (or nominal), etc.

Note.

i. QC for OEC can be either *bigger is better* or *smaller is better*. All criteria with nominal is best QC reduces to smaller is better when reading is transformed to the deviation from the nominal.
ii. When OEC is formulated correctly, sample reading with all readings as worst/best value produces 0/100 depending on the QC selected for the OEC.
iii. When relative weights are used as 0 -100, the OEC values are restricted to the same range (0 – 100)

X = Reading/evaluation of sample (for the criterion concerned)
X_L = Largest value of reading, X_S = Smallest value of reading,
X_N = Nominal/Target value for criteria with nominal QC.
B = Bigger is better, S = Smaller is better, QC = Quality characteristics

OEC = Sum of all criteria contributions

Contribution from criteria when QC is the same:

$$\left[\frac{(X - X_S)}{(X_L - X_S)}\right] \times W_T \qquad \textit{QC = B \& S}$$

$$\left[\frac{|(X - X_N)|}{|(X_L - X_N)|}\right] \times W_T \qquad \textit{QC = Nominal}$$

Contribution from a criterion when its QC needs changed:

$$\left[1 - \frac{|(X - X_{S/N})|}{|(X_L - X_{S/N})|}\right] \times W_T \qquad \textit{QC = All cases}$$

Example 1. Evaluation of Weighted Grades

		Evaluation	Criteria	Description			
#	Criteria Description	Worst Grade	Best Grade	Quality Characteristics	Relative Weighting	Sample #1 Evaluation	Sample #2 Evaluation
1	Math	0	4	Bigger	40%	3.2	3.7
2	Science	0	4	Bigger	35%	3.6	3.3
3	History	0	4	Bigger	25%	2.5	3.1
					OEC =	**79.125**	**87.0**

For Sample # 1

$$OEC = \left[\frac{(3.2 - 0)}{(4 - 0)} \right] \times 40 + \left[\frac{(3.6 - 0))}{(4 - 0)} \right] \times 35 + \left[\frac{(2.5 - 0)}{(4 - 0)} \right] \times 25$$

$= 32 + 31.5 + 15.625 =$ **79.125** *(QC = Bigger is better)*

For Sample # 2

$$OEC = \left[\frac{3.7}{(4 - 0)} \right] \times 40 + \left[\frac{(3.5)}{(4 - 0)} \right] \times 35 + \left[\frac{(3.1)}{(4 - 0)} \right] \times 25$$

$= 37 + 30.625 + 19.375 =$ **87.0** *(QC = Bigger is better)*

Example 2. Evaluation of Cake

		Evaluation	Criteria	Description			
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Sample #1 Evaluation	Sample #2 Evaluation
1	Taste	0	12	Bigger	55%	9	7
2	Moistness	25 gm	40 gm	Nominal	20%	34.9	45
3	Consistency	8	2	Smaller	25%	5	3
					OEC =	**66.95**	**66.30**

For Sample # 1

$$OEC = \left[\frac{(9-0)}{(12-0)} \right] \times 55 + \left[1 - \frac{(40 - 34.9)}{(40-25)} \right] \times 20 + \left[1 - \frac{(5-2)}{(8-2)} \right] \times 25$$

$$= 41.25 + 13.2 + 12.5 = \mathbf{66.95}$$ *(QC = Bigger is better)*

For Sample # 2

$$OEC = \left[\frac{7}{(12-0)} \right] \times 55 + \left[1 - \frac{(45 - 40)}{(40-25)} \right] \times 20 + \left[1 - \frac{(3-2)}{(8-2)} \right] \times 25$$

$$= 32.08 + 13.34 + 20.88 = \mathbf{66.30}$$ *(QC = Bigger is better)*

Example 3: Evaluation of Athletic Skills

		Evaluation	Criteria	Description			
#	Criteria Description	Worst Scores	Best Scores	Quality Characteristics	Relative Weighting	Player #1 Scores	Player #2 Scores
1	Tennis (Rank)	50	1	Smaller	20%	32	28
2	Basket Ball	0	120	Bigger	30%	65	44
3	Golf	90	32	Smaller	50%	39	34
					OEC =	**32.43**	**31.74**

For Sample # 1

$$OEC = \left[\frac{32 - 1}{(50 - 1)} \right] x\ 20 + \left[1 - \frac{65 - 0}{(120 - 0)} \right] x\ 30 + \left[\frac{39 - 32}{(90 - 32)} \right] x\ 50$$

$$= 12.65 + 13.75 + 6.03 = \mathbf{32.43}$$ (QC = Smaller is better)

For Sample # 2

$$OEC = \left[\frac{28 - 1}{(50 - 1)} \right] x\ 20 + \left[1 - \frac{(44 - 0)}{(120 - 0)} \right] x\ 30 + \left[\frac{(34 - 32)}{(90 - 32)} \right] x\ 50$$

$$= 11.02 + 19 + 1.72 = \mathbf{31.74}$$ (QC = Smaller is better)

Example 4: Comparison of Hospital Emergency Room Performance

		Evaluation	Criteria	Description			
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Hospital 1 Evaluation	Hospital 2 Evaluation
1	Evaluation	0	6	Bigger	60%	4.5	4.7
2	Complaints	50	0	Smaller	25%	22	16
3	Avg. cost	7 million$	3million$	Smaller	15%	4.8	6.5
					OEC =	**67.25**	**66.25**

For Sample # 1

$$OEC = \left[\frac{4.5 - 0}{(6 - 0)} \right] \times 60 + \left[1 - \frac{(22 - 0)}{(50 - 0)} \right] \times 25 + \left[1 - \frac{(4.8 - 3)}{(7 - 3)} \right] \times 15$$

$$= 45 + 14 + 8.25 = \mathbf{67.25}$$ *(QC = Bigger is better)*

For Sample # 2

$$OEC = \left[\frac{4.7}{(6 - 0)} \right] \times 60 + \left[1 - \frac{(16 - 0)}{(50 - 0)} \right] \times 25 + \left[1 - \frac{(6.5 - 3)}{(7 - 3)} \right] \times 15$$

$$= 47.0 + 17.0 + 2.25 = \mathbf{66.25}$$ *(QC = Bigger is better)*

Case Study I: Comparison of Vehicle Performance

New vehicles of the same class were evaluated by a consumer group based on four different criteria: Fuel economy, Cost of purchase (dealer's cost), Fault rating (published data from previous year), and Warranty. The ranges of values in each criterion were determined my market survey and focus group studies and are as shown in the table below. Among the top contenders were vehicle 1 and vehicle 2 whose data are shown in the right two columns of the table. Objectively speaking, which of the two vehicles is better?

		Evaluation	Criteria	Description		(Case Study	I)
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Vehicle 1 Evaluation	Vehicle 2 Evaluation
1	Fuel Ecnmy.	25mpg	30mpg	Bigger	15%	27	30
2	Cost- New	28k$	23k$	Smaller	32%	24.5	26.0
3	Fault Rating	8	4	Smaller	25%	7	6
4	Warranty(x1000)	50 m	80m	Bigger	28%	80	70
					OEC =	**62.65**	**58.97**

For Sample # 1

$$OEC = \left[\frac{27 - 25}{(30 - 25)}\right] \times 15 + \left[1 - \frac{(24.5 - 23)}{(28 - 23)}\right] \times 32 + \left[1 - \frac{(7 - 4)}{(8 - 4)}\right] \times 25$$

$$+ [(80 -50)/(80-50)]\mathrm{x}\ \mathbf{28}$$

$= 6.0 + 22.40 + 6.25 \quad 28.0 =$ **62.65** *(QC = Bigger is better)*

For Sample # 2

$$OEC = \left[\frac{(30 - 25)}{(30 - 25)}\right] \times 15 + \left[1 - \frac{(26 - 23)}{(28 - 23)}\right] \times 32 + \left[1 - \frac{(6 - 4)}{(8 - 4)}\right] \times 25$$

$$+ [(70-50)/(80-50)]\mathrm{x}\ \mathbf{28}$$

$= 15.0 + 12.80 + 12.50 + 18.66 =$ **58.97** *(QC = Bigger is better)*

Solution: Vehicle 1 is slightly better than vehicle 2 (62.65 vs. 58.97)

Case Study II: Automobile Dealer Performance Study

A large manufacturer of automobile used three most important criteria to determine the best dealer in the Metro Detroit area. The criteria used were: Sales volume, Warranty repair cost per vehicle (cost of repair per vehicle charged to manufacturer), and Quality of Work Life (QWL) of employees (determined by employee survey/feedback in a scale of 0 to 10). The relative importance (weight) was determined by manufacturer's dealer support organization before the commencement of the study.

		Evaluation	Criteria	Description		(Case Study	II)
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Dealer 1 Evaluation	Dealer 2 Evaluation
1	Sales Volume	600	900	Bigger	50%	800	875
2	Warr. $/veh	230	650	Smaller	30%	275	450
3	QWL-Empl.	0	10	Bigger	20%	8	6
					OEC =	**76.12**	**72.12**

For Sample # 1

$$OEC = \left[\frac{(8-6)}{(9-6)}\right] \times 50 + \left[1 - \frac{(275-230)}{(650-230)}\right] \times 30 + \left[\frac{(8-0)}{(10-0)}\right] \times 20$$

$$= 33.33 + 26.78 + 16.00 = \mathbf{76.12}$$ *(QC = Bigger is better)*

For Sample # 2

$$OEC = \left[\frac{(8.75-6)}{(9-6)}\right] \times 50 + \left[1 - \frac{(450-230)}{(650-230)}\right] \times 30 + \left[\frac{(6-0)}{(10-0)}\right] \times 20$$

$$= 45.83 + 14.29 + 12.00 = \mathbf{72.12}$$ *(QC = Bigger is better)*

Case Study III: College Admission Qualification Review

Admission to law school in one of Michigan's state funded school is decided by applicant's achievements in four different areas. The school traditionally has put 40% importance on student grade and 25% on the highest SAT score. The applicant's participation in six extracurricular activities is considered a plus and assigned 20% weight. The remaining 15% weighting is placed on the involvement (evaluated in a scale of 0 – 10) in community activities like volunteering time for social and religious organizations. The ranges of student's achievements are as shown in the table. Which among the two applicants shown in the table is a better prospect?

		Evaluation	Criteria	Description		(Case Study	III)
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Applicant 1 Evaluation	Applicant 2 Evaluation
1	SAT Score	0	1600	Bigger	25%	1450	1530
2	Grade (GPA)	0	4	Bigger	40%	3.6	3.4
3	Extra Act.	0	6	Bigger	20%	3	5
4	Comm. Serv.	0	10	Bigger	15	8	6
					OEC =	**80.66**	**83.57**

For Sample # 1

OEC = (1450/1600) x 25 + (3.6/4) x 40 + (3/6) x 20 + (8/10) x 15

= 22.66 + 36.00 + 10.00 + 12.00 = **80.66** *(QC =Bigger is better)*

For Sample # 2

OEC = (1530/1600) x 25 + (3.4/4) x 40 + (5/6) x 20 + (6/10) x 15

= 23.90 + 34.00 + 16.67 + 9.00 = **83.57** *(QC = Bigger is better)*

Case Study IV: Production and Delivery Evaluation

An automotive supplier used OEC index to compare production and delivery of its two critical parts. For comparative study, parts rejected, missed delivery dates, warranty cost data was compared. The ranges of values for each criterion and data for the two parts are shown in the table below. Which among the two parts has more room for improvement?

		Evaluation	Criteria	Description		(Case Study	IV)
#	Criteria Description	Worst Reading	Best Reading	Quality Characteristics	Relative Weighting	Part# 1 Evaluation	Part# 2 Evaluation
1	Rejects	5%	0	Smaller	60%	3	2
2	Missed Delivery	6	0	Smaller	25%	4	3
3	Warranty Cost$ Per part	3.50	0	Smaller	15%	1.5	2.2
					OEC =	**59.10**	**45.93**

For Sample # 1

$$OEC = (3/5) \times 60 + (4/6) \times 25 + (1.5/3.5) \times 15$$

$= 36.00 + 16.67 + 6.43 =$ **59.10** *(QC = Smaller is better)*

For Sample # 2

$$OEC = (2/5) \times 60 + (3/6) \times 25 + (2.2/3.5) \times 15$$

$= 24.00 + 12.50 + 9.43 =$ **45.93** *(QC = is better)*

6. Application areas

OEC can be an effective tool to formally evaluate performances of products, process, or of any service organization. It can be used in any structured studies, such as design of experiments that involve multiple factors affecting numerous objectives.

Areas of Application

- Evaluation of customer support
- Product evaluation
- Evaluation of customer survey and feedback
- Evaluation of group/departmental activities
- Analysis of results, DOE, FMEA, etc.

Method of Application

a. Identify what you want to evaluate form team
b. Hold team meeting and determine criteria of evaluations (allow sufficient time and record all ideas suggested)
c. Selected important criteria and determine their expected range of readings (All evaluations must be in numerical terms).
d. Determine relative weights of the criteria based on group consensus.
e. Complete evaluation criteria description table.
f. Proceed to formulate Overall Evaluation Criterion (OEC).

References

(The OEC concept was introduced by R. Roy in 1987)

i. Design of Experiments Using the Taguchi Approach : 16 Steps to Product and Process Improvement by Ranjit K. Roy. Hardcover (January 2001, John Wiley & Sons)

ii. A PRIMER ON THE TAGUCHI METHOD by R. Roy.

iii. OEC in Nutek Web http://www.rkroy.com/wp-oec.html

iv. Qualitek-4 software, http://www.rkroy.com/wp-q4w.html

Multiple Criteria of Evaluations for Designed Experiments

by

Ranjit K. Roy, Ph.D.,P.E.

(Published in International Quality Congress, Major Cities in India, January 1999.)

Proper measurement and evaluation of performance is key to comparing performances of products and processes. When there is only one objective, carefully defined quantitative evaluation most often serves the purpose. However, when the product or process under study is to satisfy multiple objectives, performances can be scientifically compared only when the individual criteria of evaluations are combined into a single number. This short paper deals with a method of handling evaluations of multiple objectives together by combining them into an Overall Evaluation Criteria **(OEC).**

In engineering and scientific applications, measurements and evaluations of performance are everyday affairs. Although there are situations where measured performances are expressed in terms of attributes such as Good, Poor, Acceptable, Deficient, etc., most evaluations can be expressed in terms of numerical quantities (instead of Good and Bad, use 10 - 0). When these performance evaluations are expressed in numbers, they can be conveniently compared to select the preferred candidate. The task of selecting the best product, a better machine, a taller building, a champion athlete, etc. is much simpler when there is only one objective with performance measured in terms of a single number. Consider a product such as a 9 Volt Transistor Battery whose functional life expressed in terms of hours is the only characteristic of concern. Given two batteries: Brand A (20 hours) and Brand B (22.5 hours), it is easy to determine which one is preferable. Now suppose that you are not only concerned about the functional life, but also the unit costs which are: $ 1.25 for Brand A and $1.45 for Brand B. The decision about which brand of the Battery is better is no longer straightforward.

Multiple performance objectives (or goals) are quite frequent in the industrial arena. A rational means of combining various performances evaluated by different units of measurement is essential for comparing one product performance or process output with another. In experimental studies utilizing the Design of Experiments (DOE) technique, performances of a set of planned experiments are compared to determine the influence of the factors and the combination of the factor levels that produce the most desirable performance. In this case the presence of multiple objectives poses a challenge for analysis of results. Inability to treat multiple criteria of evaluations (measure of multiple performance objectives) often renders some planned experiments ineffective.

Combining multiple criteria of evaluations into a single number is quite common practice in academic institutions and sporting events. Consider the method of expressing a Grade Point Average (GPA, a single number) as an indicator of student's academic performance. The GPA is simply determined by averaging the GPA of all courses (such as scores in Math, Physics, or Chemistry - individual criteria evaluations) which the student achieves. Another example is a Figure Skating Competition where all performers are rated in a scale of 0 to 6. The performer who receives 5.92 wins over another whose score is 5.89. How do the judges come up with these scores? People judging such events follow and evaluate each performer in an agreed upon list of items (criteria of evaluations) such as style, music, height of jump, stability of landing, etc. Perhaps each item is scored in a scale of 0 - 6, then the average scores of all judges are averaged to come up with the final scores.

If academic performances and athletic abilities can be evaluated by multiple criteria and are expressed in terms of a single number, then why isn't it commonly done in engineering and science? There are no good reasons why it should not be. For a slight extra effort in data reduction, multiple criteria can be easily incorporated in most experimental data analysis scheme.

To understand the extra work necessary, let us examine how scientific evaluation differs from those of student achievement or from an athletic event. In academic as well as athletic performances, all individual evaluations are compiled in the same way, say 0 - 4 (in case of student's grade, there are no units). They also carry the same Quality Characteristic (QC) or the sense of desirability (the higher score the better) and the same Relative Weights (level of importance) for all. Individual evaluations (like the grades of individual courses) can be simply added as long as their (a) units of measurement, (b) sense of desirability, and (c) relative weight (importance) are same for all courses (criteria). Unfortunately, in most engineering and scientific evaluations, the individual criteria are likely to have different units of measurement, Quality Characteristic, and relative weights. Therefore, methods specific to the application, and that which overcomes the difficulties posed by differences in the criteria of evaluations, must be devised.

(a) Units of Measurements - Unlike GPA or Figure Skating, the criteria of evaluations in engineering and science, generally have different units of measurements. For example, in an effort to select a better automobile, the selection criteria may consist of: fuel efficiency measured in Miles/Gallon, engine output measured in Horsepower, reliability measured as Defects/1000, etc. When the units of measurements for the criteria are different, they cannot be combined easily. To better understand these difficulties, consider a situation where we are to evaluate two industrial pumps of comparable performances (shown below). Based on 60% priority on higher discharge pressure and 40% on lower operating noise, which pump would we select?

Table 1. Performance of Two Brands of Industrial Pumps

Evaluation Criteria	Rel. Weighting	Evaluation Criteria	Pump B
Discharge Pressure	60%	160 psi	140 psi
Operating Noise	40%	90 Decibels	85 Decibels
Totals =	--	250 (?)	225 (?)

Pump A delivers more pressure, but is noisier. Pump B has a little lower pressure, but is quieter. What can we do with the evaluation numbers? Could we add them? If we were to add them what units will the resulting number have? Would the totals be of use? Is Pump A with 250 total better than Pump B?

Obviously, addition of numbers (evaluations) with different units of measurements is not permissible. If such numbers are added, the total serves no useful purpose, as we have no units to assign, nor do we know whether bigger or smaller value is better. If the evaluations were to be added, they must first be made dimensionless (normalized). This can be easily done by dividing all evaluations (such as 160 psi, 140 psi) of a criteria by a fixed number (such as 200 psi), such that the resulting number is a unitless fraction.

(b) Quality Characteristic (QC) - Just because two numbers have the same or no units, they may not necessarily be meaningfully added. Consider the following two players and attempt to determine who is better.

Table 2 Golf and Basketball Scores of Two players

Criteria	Rel. Weighting	Player 1	Player 2	QC
Golf (9 holes)	50%	42	52	Smaller
Basketball	50%	28	18	Bigger
Total Score =	--	70	70	--

Observe that the total of scores for Player 1 is (42 + 28 =) 70 and the same for Player 2 is also (52+18 =) 70. Are these two players of equal caliber? Are the additions of the scores meaningful and logical? Unfortunately, the total of scores do not reflect the degree by which Player 1 is superior over Player 2 (score of 42 in Golf is better over 52 and score of 28 in basketball is better than 18). The total scores are meaningful only when the QC's of both criteria are made the same before they are added together.

One way to combine the two scores is to first change the QC of the Golf score by subtracting from a fixed number, say 100, then adding it to the score of the Basketball. The new total score then becomes:

Overall score for Player 1 =30+ (100- 45) =85x0.50= 42.5
Overall score for Player 2 =20+ (100- 55) = 65x0.50= 32.5

The overall scores indicate the relative merit of the players. Player 1 having the score of 42.5 is a better sportsman compared to Player 2 who has a score of 32.5.

(c) Relative Weight - In formulating the GPA, all courses for the student is weighted the same. This is generally not the case in scientific studies. For the two Players above, the skills in Golf and Basketball were weighted equally. Thus, the relative weight did not influence the judgment about their ability. If the relative weights are not the same for all criteria, contribution from the individual criteria must be multiplied by the respective relative weights. For example, if Golf had a relative weight of 40%, and Basketball had 60%, the computation for the overall scores must reflect the influence of the relative weight as shown below.

Overall score for Player 1 = 30 x 0.40 + (100 - 45) x 0.60 = 45
Overall score for Player 2 = 20 x 0.40 + (100 - 55) x0.60 = 35

The Relative Weight is a subjective number assigned to each individual criteria of evaluation. Generally it is determined during the experiment planning session by the team consensus, and is assigned such that the total of all weights is 100 (set arbitrarily).

Thus, when the concerns are addressed, criteria of evaluations can be combined into a single number as described using the following application example.

An Example Application - A group of process engineers and researchers involved in manufacturing baked food products planned an experiment to determine the "best" recipe for one of their current brand of cakes. Surveys showed that the "best" cake is judged on taste, moistness, and smoothness rated by customers. The traditional approach has been to decide the recipe based on one criterion (i.e., Taste) at a time. Experience, however, has shown that when the recipe is optimized based on one criterion; subsequent analyses using other criteria do not necessarily produce the same recipe. When the ingredients differ, selection of a compromised final recipe becomes a difficult task. Arbitrary or subjectively compromised recipes have not brought the desired customer satisfaction. The group therefore decided to follow a path of consensus decision, and carefully devise a scientific scheme to incorporate all criteria of evaluations simultaneously into the analysis process.

In the planning session convened for the Cake Baking Experiment and from subsequent reviews of experimental data, the applicable Evaluation Criteria and their characteristics as shown in Table 3 below were identified. Taste, being a subjective criterion, was to be evaluated using a number between 0 and 12 and 12 being assigned to the best tasting cake. The Moistness was to be measured by weighing a standard size cake and by noting its weight in grams. It was the consensus, that a weight of about 40 grams represents the most desirable moistness, and indicates that its Quality Characteristic is of Nominal type. In this evaluation, both results above and below the nominal are considered equally undesirable. Smoothness was measured by counting the number of voids in the cake, which made this evaluation of type Smaller is better (QC). The relative weights were assigned such that the total was 100. The notations X1, X2, & X3 as shown next to the criteria description in Table 3, are used to represent the evaluations of any arbitrary sample cake.

Table 3. Evaluation Criteria for Cake Baking Experiments

Criteria Description	Worst Evaluation	Best Evaluation	Quality Characteristic(QC)	Rel. Weighting
Taste(x1)	0	12	Bigger is better	55
Moistness(x2)	25	40	Nominal	20
Smoothness(x3)	8	2	Smaller is better	25

There were three samples tested (cakes baked) in each of the eight trial conditions. The performance Evaluations for the three samples in Trial # 1 are as shown below. Note that each sample is evaluated under the three criteria of evaluations which are combined into a single number (OEC = 66 for sample 1) for each sample, which represents the performance of the cake.

Table 4. Trial# 1 Evaluations

Criteria	Sample 1	Sample 2	Sample 3
Taste	9	8	7
Moistness	34.19	33	34
Smoothness	5	4	4.5
OEC =	66.00	64	58.67

The individual sample evaluations were combined into a single number, called the Overall Evaluation Criteria (OEC), by appropriate Normalization. The term Normalization would include the process of reducing the evaluations to dimensionless quantities, aligning their Quality Characteristics to conform to a common direction (Bigger or smaller), and allowing each criteria to contribute in proportion of their Relative Weight. The OEC equation appropriate for the Cake Baking project is as shown below.

$$OEC = \frac{|(x1 - 0)|}{|(12 - 0)|} \times 55 + \left[1 - \frac{|(x2 - 40)|}{|(40 - 25)|}\right] \times 20 + \left[1 - \frac{|(x3 - 2)|}{|(8 - 2)|}\right] \times 25$$

The contribution of each criteria is turned into fractions (a dimensionless quantity) by dividing the evaluation by a fixed number such as the difference between Best and the Worst among all the respective sample evaluations (12 - 0 for Taste, see Table 3). The numerator represents the evaluation reduced by smaller magnitude of the Worst or the Best Evaluations in case of Bigger and Smaller QC's and by the Nominal value in case of Nominal QC. The contributions of the individual criteria are then multiplied by their respective Relative Weights (55, 20, etc.). The Relative Weights which are used as a fraction of 100, assures the OEC values to fall within 0 - 100.

Since Criteria 1 has the highest Relative Weight, all other criteria are aligned to have a Bigger QC. In the case of a Nominal QC, as it is the case for Moistness (second term in the equation above), the evaluation is first reduced to deviation from the nominal value (X2 - nominal value). The evaluation reduced to deviation naturally turns to Smaller QC. The contributions from the Smoothness and Moistness, both of which now have Smaller QC, are aligned with Bigger QC by subtracting it from 1. An example calculation of OEC using the evaluations of Sample 1 of Trial # 1 (see Table 4) is shown below.

Sample calculations: Trial 1, Sample 1 (x1 = 9, x2 = 34.19, x3 = 5) OEC = 9 x 55 / 12 + (1 - (40 - 34.19)/15) x 20 + (1 - (5 - 2)/6) x 25 = 41.25 + 12.25 + 12.5 = 66 (shown in Table 4)

Similarly, the OEC's for all other sample cakes for Trial # 1 (OEC = 64 for Sample 2 and OEC = 58.67 for Sample 3) and other Trials of the experiment are calculated. The OEC values are considered as the "Results" for the purposes of the analysis of the designed experiments.

The concept OEC was first published by the author in the reference text in 1989. Since then it has been successfully utilized in numerous industrial experiments, particularly those that followed the Taguchi Approach of experimental designs. The OEC scheme has been found to work well for all kinds of experimental studies regardless of whether it utilizes designed experiments.

Author: Ranjit K. Roy

References:

1. A Primer on The Taguchi Method by Ranjit K. Roy, Society of manufacturing Engineers
2. Qualitek-4 Software for Automatic Design and Analysis of Taguchi Experiments