

Combining Computer-Aided Statistical Design of Experiments with “Traditional” Design of Experiments to Meet EC2000 Criteria

David R. Mills¹

Abstract

New ABET criteria require that chemical engineering students be able to “design and conduct experiments”. Experimental design may be interpreted as either developing methods and procedures to achieve experimental goals, or as statistical “Design of Experiments” (DOE). Both approaches provide students with important skills. The first approach allows students to exercise creativity, and to learn that methods/procedures must be well thought out to achieve goals in a real process. The latter approach provides statistical skills widely used for industrial quality control and improvement, as well as in laboratory research. This is an area in which chemical engineering curriculum has typically lagged behind industrial practice. In our unit operations lab course we give students a project in which they must come up with their *own* factorial experiments. In these projects, students “design” methods to achieve their experimental goals and also incorporate DOE into procedures and analysis of results. Students acquire skills deriving from *both* definitions of experimental design using our approach. Additionally, they generally report having “more fun” doing their own projects, and seem to approach them with more enthusiasm and hard work than other lab experiments, thus enhancing learning. Design-Expert[®] software is used for designing and analyzing experiments, and to optimize conditions. This software assists instruction, is user friendly and allows students to choose from many possible designs. The “expert” system guides students through the interpretation of ANOVA output, allowing them to learn DOE quickly and explore options on their own. Students have demonstrated the ability extend concepts and use the software for other experiments after using the software once. Feedback from students on end-of-course surveys has been positive. In addition to achieving beneficial learning outcomes, we think that our dual approach is the best way to satisfy EC2000 criteria. To that end, project grades are partly determined using a rubric that covers elements of both definitions of “design of experiments” and doubles as an ABET assessment form. A case study of a student experiment is given.

“Design of Experiments”: What Does it Mean?

EC 2000 Criteria [ABET,1] states that students “have the ability to design and conduct experiments”. However, the definition of “design of experiments” is not given and we are left to interpret this term ourselves. Traditionally, the term has meant writing experimental procedures that insure the experimental goal is met. Many educators seem to adhere to this definition. [Lauterbach, 5] [Machias-Menton, 6] [McIntosh, 7] [Middleberg, 8] [Rosenthal, 12] [Tawney, 13] Some of these articles appear in the ASEE publication, “Chemical Engineering Education”, and pertain specifically to the chemical engineering “unit ops” courses at various universities. [Lauterbach,5] [Machias-Menton, 6] [Middleberg, 8] Other authors [Dorland, 3] [Munson-McGee, 10] interpret “design of experiments” as formal “statistical Design of Experiments” (hereafter referred to as “DOE”- we will refer to the more traditional definition of design of experiments as “doe”) in their application to the unit operation courses. Others [Machias-Menton, 6] [Hudgins, 4] at least make the case that related statistical tools need more emphasis in chemical engineering curriculum. Munsen-McGee [10] takes the DOE approach to the limit, outlining

¹ Department of Chemical Engineering, Auburn University, Auburn, AL 36849 (334-844-2015)

a unit operations course where fractional factorial, Graeco-Latin squares, blocking, nested, and mixtures designs are all investigated twice during the course and are apparently applied to all experiments in the unit ops course.

The DOE interpretation appears mostly in more current literature. This is expected, as statistical methods such as DOE and Statistical Process Control (SPC) have gained industrial importance in recent years. The demand for engineers trained in these areas is evident from a simple internet search using the term “DOE”. Such a search returned over 80 pages of results including frequent entries for consulting services, training programs, and textbooks aimed at practicing professionals. An additional motivation for us is that a member of our Alumni Council has spoken emphatically of the need to include this material in the curriculum. Clearly, there is a need for engineering students to have increased exposure to “real-world” data that includes uncertainty, and to be able to design experiments and interpret data as part of their undergraduate education. The use of “expert-system” software for DOE is a novel feature of our approach that is not found in the literature.

On the other hand, we feel it is still necessary for students to become accustomed to writing their own effective procedures (doe) in “open ended” experiments. We agree with others that this approach can foster higher-order learning skills as defined by Bloom [2], [Miller, 9]. We do not feel that it is necessary to rely *entirely* on statistical methods for experimental design. Over reliance on statistical DOE and purely empirical models may hinder other goals of enhancing the understanding of transport processes and unit operations. For these reasons we still grade and evaluate experimental planning, execution of procedures, and appropriate use of equipment and methods as “experimental design.” In short, we believe a combination of DOE and “doe” to be the best approach.

A Combined Approach to Design of Experiments

Curriculum and Lab Project

Our “Unit Ops” curriculum is currently organized into two, two semester credit hour courses with one hour lecture and three hours of laboratory per week. The first course focuses on momentum and heat transfer, while the second focuses more on mass transfer, separations, and reactor operations. In the first course we confine statistical analysis to propagation of error, statistical analysis of uncertainties and linear regression. The textbook, “An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements”, by John R. Taylor (University Science Books, Sausalito, CA) is an excellent text we use for this purpose. In the second Unit Ops course we introduce DOE and analysis of variance (ANOVA). However, we still maintain a strong focus on traditional “doe” skills.

We chose to use Stat-Ease’s Design-Expert[®] software for teaching and applying DOE because it is easy to use, powerful, and provides “expert system” guidance throughout the design and analysis process. The developers of the software are both chemical engineers and thus provide numerous chemical engineering examples in an extensive on-line tutorial. The software was inexpensive (about \$200 for a lab site license to use during instruction) and students can download fully functional trial editions from the company web site. Stat-Ease founder Pat Whitcomb came to our department and gave a seminar to help us get started. The use of “expert system” software to teach DOE to engineers at the university level is a relatively new idea. However, we were interested to find that at the time of our seminar, Stat-Ease had been contacted by chemical engineering departments at South Dakota School of Mines and the University of Minnesota and asked to present future seminars there. [Whitcomb, 14] The fact that other departments are independently starting to think along the same lines makes us even more confident that our approach to DOE is a good one.

Students chose their own “design of experiments” projects. Those projects with specific chemical engineering applications were encouraged but not required; we mainly wanted students to pick a subject of interest to them. An additional benefit to having students pick unique projects was that it eliminated the potential problem of students copying lab reports. It also eliminated the need to have duplicate lab apparatus, so we were able to cover the DOE material in lecture while all students were working on DOE lab projects. Four lectures were dedicated to DOE and ANOVA. Students either scheduled lab time to perform experiments or in many cases were able to do projects at home. Working at home may have encouraged more creativity and “trial and error” since students could explore options without the fear of making mistakes in front of the TA or instructor. Our approach covers all the “design of experiments” bases: students must come up with a goal, experiment and procedures to reach the goal (doe), and

use DOE to formally design the experiment. As a starting point, students were asked to focus on a factorial experiment of some type. Students came up with an idea for an experiment and met with the instructor. After approval of the project, another meeting was scheduled in which students brought in experimental designs. These included goals, equipment, and procedures, along with a factorial design and randomized run sheet generated in Design-Expert[®]. Students had three weeks to complete their projects and turn in a written report. The report included statistical analysis (ANOVA), models, and optimization where appropriate.

General Observations

Students seemed to enjoy this project more than any of the other laboratory projects. Sample report titles were “Expertly Designing Better Burgers”, “Analysis of the Effects of Stir Rate, Temperature and Crystal Size upon the Speed of Dissolution of Sodium Chloride in Water”, “Analysis of Major Cola Products: Coke vs. Pepsi”, “Effects of Various Factors on the Quality of Tea”, and “Design Experiments Correlating Factors That Affect Paper Towel Performance” and “Two Factorial Laundry Experiments”.

The software proved to be easy to use and students had few problems with the initial DOE. They were able to explore different options for their particular project such as which design to use (two factor, irregular fraction, D-optimal, etc.), whether to use full or half factorial experiments, and consider how to use blocking and replicates where appropriate. The ANOVA and related analysis were generally very good, especially given the limited statistical background of many students and the limited time for coverage of ANOVA material in lecture. The “expert system” features and help menus of the software were of great help in this area. Still, ANOVA results were sometimes over-interpreted; students often relied too much on the statistics and didn’t use enough engineering analysis to make sense of results. We hope to see improvements as we address this issue in class, add more statistics to courses taken earlier in the curriculum and add an additional lecture on DOE to the lab course.

On the “doe” front, some groups had to come up with very creative apparatus and measurement techniques to achieve experimental goals. Each project had its own challenges, and as the projects evolved students quickly came to appreciate what I often tell them: “there is no such thing as a *simple* experiment”. Projects involving human subjects and non-quantitative results generally gave the least satisfactory results. We plan to coach students not to choose those types of projects in the future. The reasons are twofold. The first is that experiments of this type are probably more appropriate to the social sciences and we do not have time to focus on topics such as ensuring double blindness, etc. Secondly, even though these experiments are informal, we decided we may be at risk for violating policies on the use of human subjects.

Quantitative Results and Documentation for ABET

Each student was responsible for writing up results and submitting an individual report. For ABET documentation, a randomly selected sample representing 50% of the reports were evaluated by the course instructor and two other faculty members. The Data Analysis/Experimental Design Assessment Rubric shown in Figure 1 was used to assess student’s ability to design and conduct experiments, and to analyze the data. Results compiled from all faculty evaluations using the rubric in Figure 1 are shown in Table 1. The rubric in was designed to be simple to use and to be applicable to either definition of “design of experiments”.



Data Analysis / Experimental Design Assessment Rubric

Course No.: _____ Date: _____
Team/Student: _____ Reviewer: _____

Topic (Weight)	Unacceptable (0)	Marginal (1)	Acceptable (2)	Exceptional (3)	Points
Effectiveness of Experimental Design and/or Procedures (2)	0 Very ineffective. Would not allow experimenters to achieve any goals.	0 Somewhat ineffective. Would allow experimenter(s) to achieve some goals.	0 Somewhat effective. Would allow experimenter(s) to achieve most goals.	0 Effective. Would allow experimenter(s) to achieve all goals.	
Execution of Procedures (1)	0 Demonstrated little or no ability to conduct experiments. Did not collect meaningful data.	0 Demonstrated some ability to conduct experiments. Collected some meaningful data.	0 Demonstrated adequate ability to conduct experiments. Collected most of the needed data.	0 Demonstrated superior ability to conduct experiments. Collected all the appropriate data.	
Statistical Methods: Error Analysis, Regression, ANOVA (2)	0 Statistical methods were completely misapplied or absent.	0 Statistical methods were attempted. Some methods were applied but with significant errors or omissions.	0 Statistical methods were attempted. Most methods were correctly applied but more could have been done with the data.	0 Statistical methods were fully and correctly applied.	
Focus of Results and Discussion (1)	0 No insight. Entirely missed the point of the experiment.	0 Little insight. Analyzed only the most basic points.	0 Adequate insight. Missed some important points.	0 Excellent insight. Results and discussion well focused.	
Interpretation of Data (2)	0 Little or no attempt to interpret data or over-interpreted data.	0 Interpreted some data correctly. Significant errors, omissions, or over-interpreted data.	0 Interpreted most data correctly. Some conclusions may be suspect or over-interpreted.	0 Data completely and appropriately interpreted. Not over-interpreted.	
OVERALL PERFORMANCE	0 Unacceptable	0 Marginal	0 Acceptable	0 Exceptional	TOTAL
POINTS REQUIRED	0-6	7-12	13-18	19-24	

Figure 1- Rubric used to compile data for assessing design/ conduction of experiments and data analysis. Each item is weighted by a factor of 1 or 2 as shown in the left most column. A score of 19-24 is considered “excellent”, 13-18 “superior”, 7-12 “marginal” and 0-6 “unacceptable”. An average score from 3 reviewers is used to measure student performance.

Report No./Group No.	Numerical Scores					Categorical Scores				
	Reviewer 1	Reviewer 2	Reviewer 3	Avg. Score	Range	Evaluator 1	Evaluator 2	Evaluator 3	Average	
1-(1)	15	6	10	10.3	9	A	U	M	M	
2-(1)	13	6	8	9.0	7	A	U	M	M	
3-(1)	13	6	10	9.7	7	A	U	M	M	
4-(2)	11	8	16	11.7	8	M	M	A	M	
5-(3)	20	18	18	18.7	2	E	A	A	A	
6-(4)	11	19	18	16.0	8	M	E	A	A	
7-(5)	19	22	19	20.0	3	E	E	E	E	
8-(4)	13	22	18	17.7	9	A	E	A	A	
9-(4)	13	15	17	15.0	4	A	A	A	A	
10-(6)	17	22	19	19.3	5	A	E	E	E	
11-(6)	17	22	16	18.3	6	A	E	A	E	
12-(3)	17	20	17	18.0	3	A	E	A	A	
Average:	14.8	15.6	16.0	15.5	5.3	A	A	A	A	
Standard Deviation:	3.2	6.0	3.6	3.6	2.7					
						Key:	E = Exceptional A = Acceptable M = Marginal U = Unacceptable			

Table 1- Summary of faculty evaluation for DOE based projects entirely of student creation.

All evaluators found the overall average student performance to be acceptable. The range of scores by different evaluators for a particular report deserves comment. Reports 1-3 were by different students in the same group performing the same experiment and show marked disagreement between evaluators. Reviewer 2 was especially critical of the choice of experiment and therefore the “effectiveness of the design and/or procedures”. The other two reviewers still disagreed somewhat, with reviewer 3 giving the report a marginal rating whereas reviewer 1 deemed the report acceptable. Reports 6 and 9 also show wide variability between reviewers, with a range of 8 and 9 respectively. Evaluator 1 was the course instructor and may have had “inside information” that influenced assessment. Averaging the scores helped, but to improve consistency, we will in the future try to have better agreement on definitions. Even so, some variation is to be expected as individual evaluators naturally have subjective preferences. It may also be desirable to meet for discussion before assigning final scores.

There was a fairly even distribution of Exceptional, Acceptable, and Marginal efforts. Although these were individual reports, each evaluator tended to grade students from the same group similarly, indicating either the importance of common elements in the rubric such as “Effectiveness of Design and/or Procedures”, or the tendencies of individuals to form groups where each member has similar abilities (students were permitted to choose their own groups), or both.

Case Study of a Student Project

This is an example of a 2-factorial experiment that students designed (doe and DOE) themselves. This group decided to investigate the effect of laundry detergent brand, temperature, and wash time on the effectiveness of stain removal. The types of stains investigated were ketchup, mud, mustard, tea, and red wine. The students were faced with several challenges in designing the methods and procedures which were brought out in meetings with the instructor. Mainly, they had to assure consistency in staining and to devise a measuring system to gauge “stain removal effectiveness”. To get around the first problem they decided to cut up a white cotton T-shirt and stain a section with the staining substance of interest. They then cut each stain into pieces to be used in the washing processes. To gauge the degree of stain removal they used a color printer to make a color comparison scale. This consisted of a base color for each type of stain and series of lighter shades of the base color numbered on a 1-10 scale. This creative approach demonstrated proficiency in one aspect of traditional design of experiment: designing methods, procedures and apparatus.

Students chose a two-factorial statistical design for this project. They were concerned with repeatability and consistency of results so decided to do three replicates of each experimental run. The replicates allowed for a better estimate of pure error. Since the students did the experiment over two days, they set up separate blocks for each day in order to avoid any “lurking” time effects. The following graphics are intended only to illustrate some of the pedagogical applications of the Design-Expert[®] software for this particular factorial experiment, and are not intended to show the full capability of the product.

Figures 2-3 are screen capture images that show the main screens used for designing the experiments. Figure 3 also shows the student’s experimental data for this project, entered after all runs were completed. Figure 4 is a half-normal plot that shows significant main effects and interactions between the effects (interestingly, in this experiment some responses showed no significant effects while some showed only main effects, and others showed interaction effects as in Figure 5).

After selecting likely significant effects, an ANOVA printout (Figure 6) is generated to verify that effects are significant. The ANOVA printout also gives model results and has automatically generated annotations that help greatly with interpretation of results. We feel that this feature is especially useful for students with very limited experience in DOE and ANOVA. Figure 5 is an example of a model graph that shows an interaction for the “wine” effect. Not shown are other options such as contour, 3-D surface, and cube plots. Also not shown are the many diagnostic plots such as various residual plots, outlier T plots, leverage, actual vs. predicted, and box plots. Optimization features are also available; students could have optimized conditions to achieve best performance for all stains in this case.

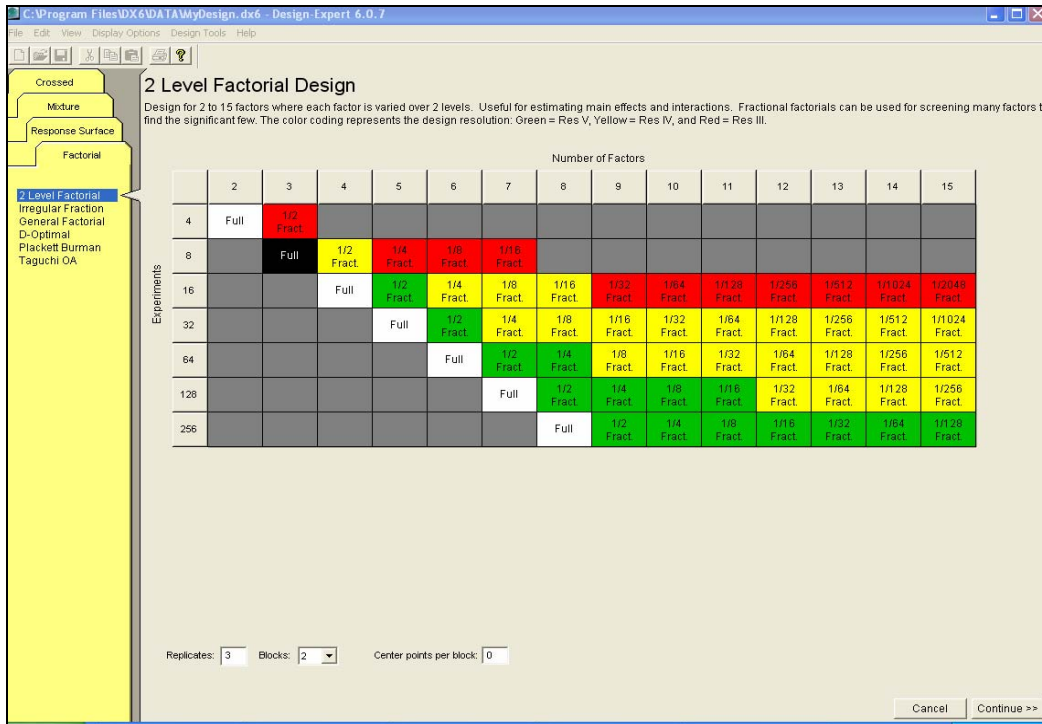


Figure 2- Design-Expert screen showing starting point for design selection. Note that a variety of design types are available on the tabs to the left of the screen, with sub categories of each design.

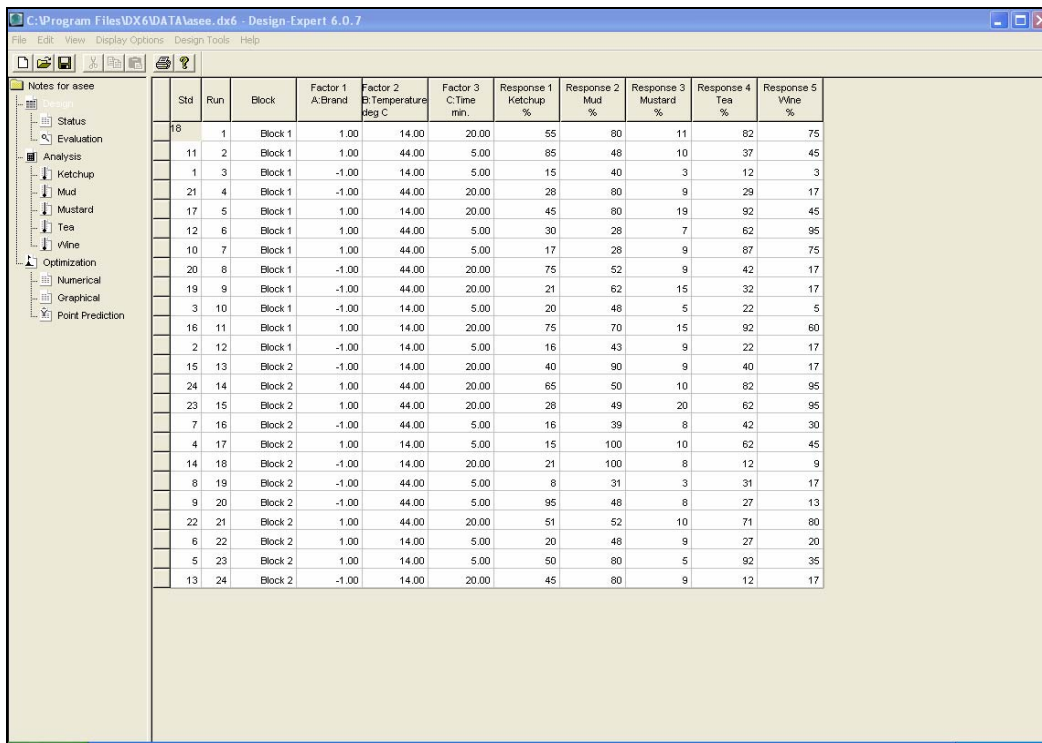


Figure 3- Design-Expert screen of design conditions and results. Columns 1-6 are computer generated. Columns 1-2 show both standard and actual run order, randomized by the software. Columns 7-11 are results that have been filled in by the experimenters with percentage stain removal for each type of stain.

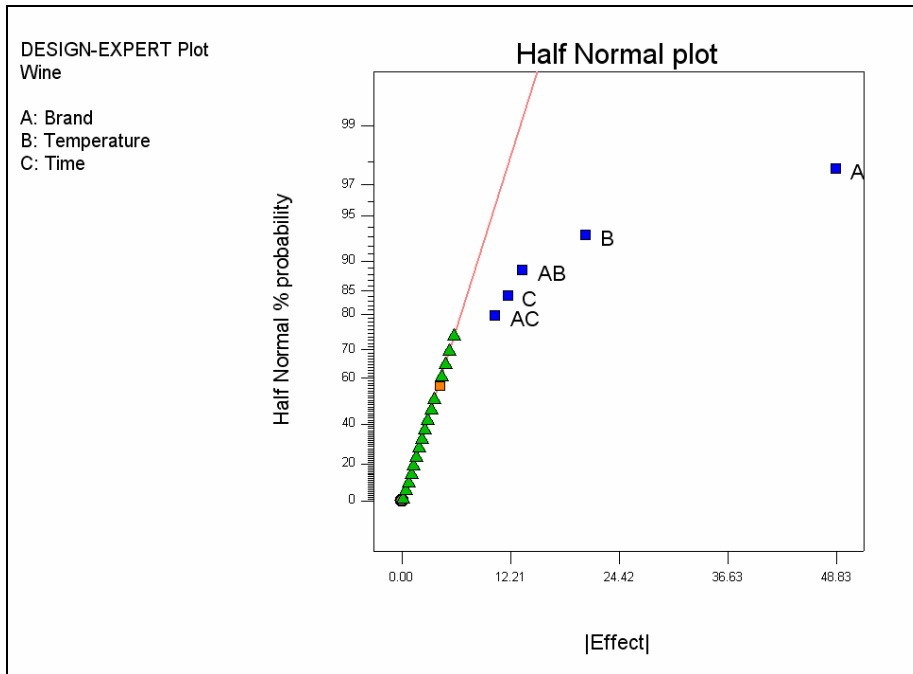


Figure 4- Half normal plot for wine stains. Here all of the main effects (A, B, C) appear to be significant, as are interactions between brand of detergent (A), and both time and temperature (C). Students have highlighted these terms by mouse click and they appear as blue squares. These effects are then modeled in subsequent procedures.

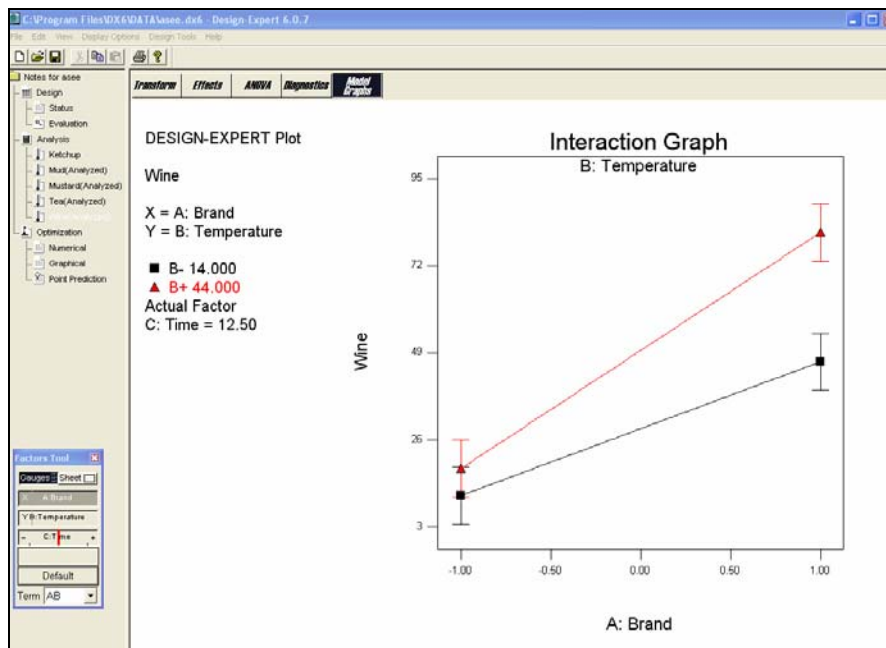


Figure 5- Various graphical options are given for data and model presentation. This plot shows the interaction effect for brand and temperature on the % stain removal for red wine. The slider at the left-bottom of the screen can be used to change the temperature or look at different factors and interactions.

Use your mouse to right click on individual cells for definitions.

Response: Wine

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	0.166667	1	0.166667			
Model	19489.83	5	3897.967	25.37609	< 0.0001	significant
A	14308.17	1	14308.17	93.14737	< 0.0001	
B	2562.667	1	2562.667	16.68318	0.0008	
C	864	1	864	5.624713	0.0298	
AB	1093.5	1	1093.5	7.118777	0.0162	
AC	661.5	1	661.5	4.306421	0.0535	
Residual	2611.333	17	153.6078			
Lack of Fit	112.6667	1	112.6667	0.721451	0.4082	not significant
Pure Error	2498.667	16	156.1667			
Cor Total	22101.33	23				

The Model F-value of 25.38 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Values of "Prob > F" less than 0.0500 indicate model terms are significant. In this case A, B, C, AB are significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model.

The "Lack of Fit F-value" of 0.72 implies the Lack of Fit is not significant relative to the pure error. There is a 40.82% chance that a "Lack of Fit F-value" this large could occur due to noise. Non-significant lack of fit is good -- we want the model to fit.

Std. Dev.	12.39386	R-Squared	0.881846
Mean	39.33333	Adj R-Squæ	0.847095
C.V.	31.50982	Pred R-Sq	0.76451
PRESS	5204.595	Adeq Preci	12.20098

The "Pred R-Squared" of 0.7645 is in reasonable agreement with the "Adj R-Squared" of 0.8471.

"Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. Your ratio of 12.201 indicates an adequate signal. This model can be used to navigate the design space.

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High	VIF
Intercept	39.33333					
Block 1	-0.083333	1				
Block 2	0.083333					
A-Brand	24.41667	1	2.529887	19.07907	29.75426	1
B-Temperature	10.33333	1	2.529887	4.995739	15.67093	1
C-Time		6	2.529887	0.662406	11.33759	1
AB	6.75	1	2.529887	1.412406	12.08759	1
AC	5.25	1	2.529887	-0.087594	10.58759	1

Wine =

9.355556

2.616667 * Brand

0.688889 * Temperature

0.8 * Time

0.45 * Brand * Temperature

0.7 * Brand * Time

Figure 6- Partial ANOVA results from Design-Expert for the “wine” effect. The actual output is more extensive, but has been edited to conserve space here. Annotations in the ANOVA are computer generated and assist students with interpretation of the data as well as reinforcing lecture material.

Carry-Over Into Other Experiments

Although the techniques and software used in the DOE module were not required in other laboratory projects, some students took the initiative to use the techniques learned and tools available for other experiments. We were encouraged to see that the four lectures and one lab project with Design-Expert[®] provided learning sufficient to allow students to make effective use of DOE. This was also evidence of them applying higher order learning skills in adapting the skills learned to another project.

An example of an experiment where students applied the DOE methods was an oxygen mass transfer lab where students characterize K_1a as a function of gassed in power and superficial velocity in a 16 L stirred tank bioreactor. Most literature correlations for K_1a take the form of Eqn. 1 where P_g is gassed in power (a function of both impeller speed and sparge rate) and v_s is superficial gas velocity in the fermenter.

$$K_1a = aP_g^b v_s^c \quad (1)$$

We traditionally asked our students to find parameters a, b and c from regression and to compare to literature correlations. However, one group of students adapted the DOE methods to develop a randomized factorial design, and found a polynomial model to compare as well. The results were interesting in that they were able to correlate K_1a using simpler factors (sparge rate and impeller speed) and still capture the interaction of the two factors using the Design-Expert[®] software. This was a good alternative to using the interaction inherent in the gassed-in power factor.

Concluding Summary

EC2000 criteria pertaining to design of experiments can be interpreted in one of two ways: 1) design of methods and procedures or 2) statistical DOE. Our approach of having students work on projects that include aspects of *both* interpretations has been favorably received by students and faculty alike. Students gain the benefits of each approach: experience employing creativity to develop well thought out methods and procedures to achieve experimental goals in a real process for the first case, and learning to employ powerful statistical DOE methods and tools in the latter case. Early data from student course surveys combined with faculty observations and grading indicates that our approach is very effective for teaching design of experiments. The fact that students are able to use the Design-Expert[®] software effectively after just one project testifies to the efficiency of learning under the somewhat novel system we have employed.

Although our program has not yet been evaluated under EC2000, we think that our dual approach is the best way to satisfy the EC2000 “design and conduct experiments” criteria. In preparation for ABET evaluation, an effective design of experiments grading rubric and ABET assessment form has been developed and presented within.

References

1. Accreditation Board for Engineering Technology (ABET) (1999) "Criteria for Accrediting Engineering Programs," <http://www.abet.org>, Baltimore, MD
2. Bloom, B.S. Ed., *Taxonomy of Educational Objective*, David McKay Co., New York (1956)
3. Dorland, Dianne, Yin, Karen K, (1997) "Teaching Statistics to ChE Students", , **31**(3), 168
4. Hudgins, R.R., Reilly, P.M. (1989), "Do Student Engineers Understand Experimental Error", *Chem. Eng. Ed.* **23**(2), 92
5. Lauterbach, J., White S., et al. (1997) "A Novel Laboratory Course on Advanced ChE Experiments", *Chem. Eng. Ed.*, **31**(4), 260
6. Machias-Machin, A., Zhang G., et al., (1990) "The Unstructured Student-Designed Research Type of Laboratory Experiment" , *Chem. Eng. Ed.*, **24**(2), 78
7. McIntosh, Thomas C., (1995), "Problem-Solving Practice: Challenging Students to design Experiments and Organize Data," *The Science Teacher.* **62**(6) 48
8. Middleberg, Anton P., (1995) "Laboratory Projects- Should Students Do Them or Design Them." *Chem. Eng. Ed.* **29**(1), 34
9. Miller, Ronald L., Ely, James F., et al., "Higher Order Thinking in the Unit Operations Laboratory", *J. Chem. Ed.* **32** (2), 146
10. Munson-McGee, Stuart H. (2000), "An Introductory ChE Laboratory incorporating EC 2000 criteria", *Chem. Eng. Ed.*, **34**(1), 80
11. Rosenthal, D., Arnold, D. (1997), "Simulation of Experimental Data, the Design of Experiments and the Analysis of Results", *J. Chem. Ed.* **54** (5), 323
12. Tawney, D.A. (1972) "The Design of Experiments and the Estimation of Experimental Errors: A necessary Preparation for project work", *Physics Education.* **7**(6), 377
13. Whitcomb, Patrick J., Stat-Ease, Inc., (2002), Personal communication

David R. Mills

Mr. Mills holds an MS in Chemical Engineering from Washington State University. He has been involved in the teaching of unit operations laboratory courses in the chemical engineering departments of the University of California at Davis, Washington State University and Auburn University. He has been responsible for major overhauls and engineering of new laboratory equipment at the latter two institutions and currently instructs unit operations laboratory courses at Auburn, and is Manager of Process Engineering Laboratories there. He is the 2003 recipient of one of the top two teaching awards in the Auburn College of Engineering for his work on the unit operations courses and facilities.