The One-Hour Design of Experiments Course

Bjarne Bergquist, Erik Lovén & Thomas Olsson
Division of Environmental & Quality Management
Luleå University of Technology, Sweden

Contact information to corresponding author:

Bjarne Bergquist

Division of Environmental & Quality Engineering
Luleå University of Technology
97187 Luleå
Sweden

email. bjarne_b@ltu.se

Abstract
How we make our propositions attractive is, of course, context dependent. This article describes how a one-hour introductory course in design of experiments (DOE) can be used to attract high school students to study science and engineering in general and quality management in particular. The course is based on a 22 factorial design inspired by the Pepsi Challenge and is evaluated in the light of the "Seven Good Principles to Achieve Good Learning." Results of the course show that it is possible to teach DOE to students in high school and at the same time provide a fun experience. Many of the statistical concepts appear to have been grasped up to the level of comprehension of effects and statistical significant effects compared to random error effects.

Keywords
Design of Experiments, Experimental design, Short course, Statistics education, Student retention

Introduction
Could the basic concept of factorial design be taught to upper secondary school pupils without prior knowledge of statistics, and could this be done within one hour with the ultimate goal to attract them into undergraduate studies in science and engineering? This paper describes such an experiment and discusses the outcome.

One might think that presumptive students would consider the topic of science and quality management is interesting without much convincing and that student retention is not a problem, but this is not so. A decreased numbers of presumptive students due to variations in nativity in combination with a large expansion of Swedish higher education positions has made student retention a major issue in Sweden. Vacant positions, especially in engineering, are one of the reasons for why two thirds of Swedish seats of learning did show a negative balance of accounts for the year 2004 (National Agency for Higher Education, 2005). To make things worse, our university, Luleå University of Technology is situated far from Swedish population centres and has a focus on engineering. Thus, student retention is an issue and requires that whatever that is done to attract students is fun, interesting, yet credible.

To interest upper secondary school pupils in university studies, Luleå University of Technology arranges short courses at campus. Each of these courses spans one day and in one of these courses, the chosen topic was quality and environmental management. Within this day, Design of Experiments (DoE) was thought appropriate to represent scientific inference and was assigned one hour of the day. Such a short course will naturally not give students a deep understanding of the subject. However, a one-hour course could be enough to inspire to learn more. As students should be attracted to study at our university and especially taking courses in quality and statistics, the short courses had to be fun. However, DoE is based on statistics and maths, words that in Sweden these days seems to repel rather than attract youths.

Purpose
The purpose of the study was to design and study outcomes of a one-hour DoE course as a means to introduce science and statistics for pupils with little or no statistical background.

Method
Since the One-Hour-DoE-Course was a first time test, there was no previous knowledge of how a course of this type could work and it will thus not be possible to conclude if the course is a better way than any other. The classroom is a complex social environment with pupils interacting in-between as well as with teachers. Action research is an established method of doing research similar situations and has been tested favourably also in classrooms (Macintyre 2000).

Inspired by the Pepsi Challenge, a 2^2-experiment was designed where the participators should rate how good five beverages tasted on a ten-point Likert scale, see Table 1.

Table 1. Experimental arrangement of original experiment and grades from one of the participants.

Beverage. No:	Cola type	Sweetener	Cola/ Sweetener interaction	Grade
4	-	-	+	*7*
2	+	-	-	*4*
1	-	+	-	*7*
3	+	+	+	*2*
5	0	0	0	*9*
Effects	*-4*	*-1*	*-1*	

The data was obtained by observation of the participants. Observation was selected over interviews so that the data collection should be as non-intrusive as possible so that the data collection should not intrude and be a repellent factor for the pupils.

During the course, the first author of the paper acted as teacher and the second and third author acted as observers. All authors were active in planning and evaluating the experiment. The two observers did after completion of the course write down what they had observed regarding participants' questions, face expressions, and other things they noticed. The observations were also reflected individually in light of the *seven good principles to achieve good learning* (Chickering and Gamson, 1987). These principles are based upon decades of

research on learning in higher education and these principles has been the base of numerous evaluation programmes of higher education. These principles are:

1. High student-faculty contact,
2. Cooperation among students,
3. Active learning,
4. Prompt feedback,
5. Time on task,
6. High expectations
7. A respect of diverse talents and ways of learning.

The two observer reports were then compared to triangulate the observations.

Theory
According to Bisgaard, (1991), any introduction to statistics given to engineers or aspiring engineers should start with a practical problem so that students would learn first-hand what statistics could do to solve the problem.

An introduction to statistics could of course be done in several different was of which the topic of DoE is one. There is little written about short introductory courses in DoE. Lawrence (1996) discussed a two hour workshop to introduce statistics by testing short term memory using a 2^3-design to first year university students. Another description of first-level, but more extensive introductory course in DoE is given by Andersson-Cook (1998).

Realization
The one-hour course was introduced to the pupils with a quick discussion of the importance of making experiments and the importance of extracting as much information as possible from performed experiments. All participating pupils were given five glasses of beverages from five different bottles. The bottles were masked and all beverages had the same temperature. Main effects were explained as the main purpose of the test, introduced by the question "What is more important to You, beverage type and if so, which Cola do You prefer, or is it the sweetener that is the most important thing for You?" A square was drawn on the blackboard and the factors of Cola type and Sweetener was drawn showing the experimental grid. Taste effects were then explained as the difference in means in tastes between the two levels of the factors, all the time referring to the cube. A fifth beverage was also described as a centre point experiment consisting of an equal mix of all beverages. The centre point was not a true centre point since factors were categorical, but this was disregarded as it was not considered important for the discussion.

Glasses were distributed to all participants and placed upon a large sheet (40 by 60 cm). The sheet had a grid of five columns and ten rows, see Figure 1. The rows indicated taste from 1 = awful to 10 = delightful and the columns were used to separate the five beverages. The participants were then instructed to take a sip of each beverage and then place it on the grid corresponding to how it tasted and then to translate the placement to taste scores on a score sheet.

Figure 1. Response grid and test situation

A score sheet of a participant was selected and used the score to describe the principles of factorial design, starting placing the grades at the cormers and centre of the square on the blackboard, see Figure 2.
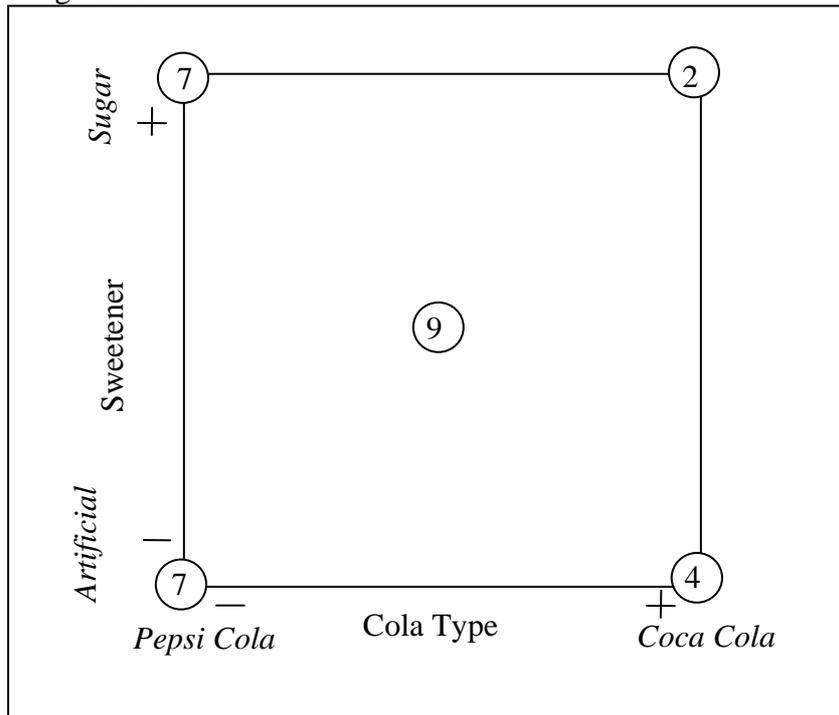

Figure 2. Square plot of experimental layout.

The square was displayed on the classroom blackboard and was used to demonstrate the effects of changing from Coke to Pepsi or if sugar really tastes better. The effects were calculated showing the principle of effects, after which all participants calculated the effects of their own tasting tests.

The interaction effect was described to the students with the following: assuming we rate sugar beverages higher than those beverages with artificial sweetener, it is not necessarily that there is an equal taste difference between Coca Cola's sugar and diet version and Pepsi's sugar and diet versions. Thus, to evaluate how we feel about artificial sweeteners, it matters if we drink Pepsi or Coke! This is an interaction effect and to find interaction effects, we must study two or multiple factors simultaneously.

Now the students were asked if these effects that they saw was due to randomness or if there actually was a difference between Pepsi or Coke, and this inspired a discussion of statistical significance. Since we did not have a replicate measure, we could not evaluate if effects were real or if they were random, so therefore the selected participant were given an additional blind test of one beverage drink, which she had to grade. After the extra trial had been unmasked, it was revealed that the extra test was graded identical to the original test score of that beverage. Since the difference between the two grades were zero, the best guess of how much variation there was in the grading was zero and therefore all effects smaller or larger than zero should be real and the plausibility of that statement was discussed.

The centre point experiment let us discuss model assumptions and violations. One assumption of two-level type factorials is that the effects are linear. The respondent we picked had rated the centre point the highest and this inspired to a lively debate. What model could be accurate if this experiment should fit? A quadratic curve that might explain the relationship was drawn on the blackboard and the meaning of the result was discussed with the participants, see Figure 3.
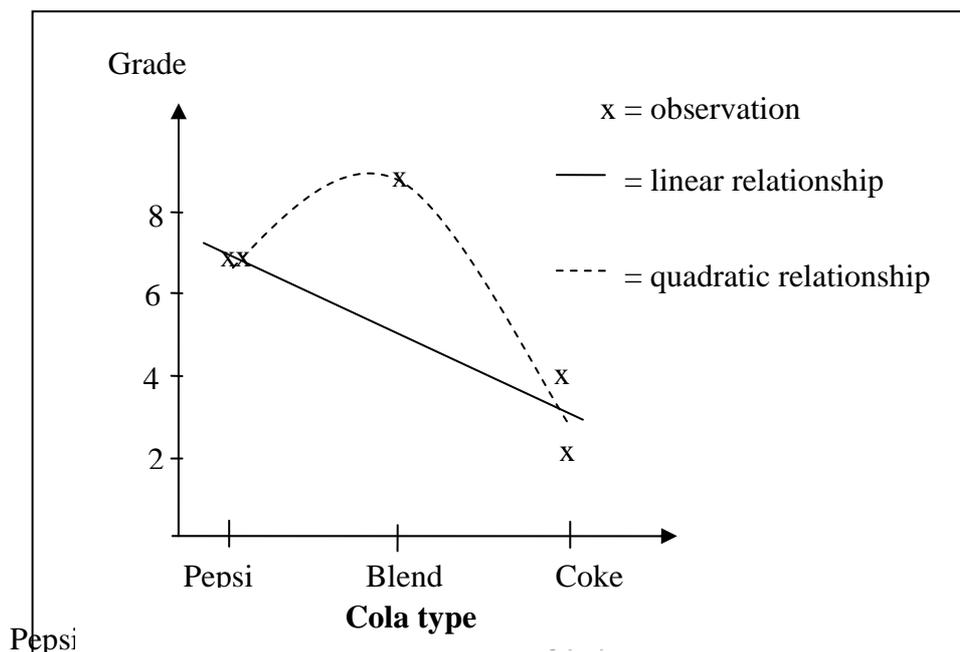

Figure 3. Possible relationship between cola type and grade.

Before the hour had passed, the experiment was again reflected to how and where experimental design could be used.

Results

The teacher as well as the observers where unite in their view that the participants had thought that the experiment was fun, and . One observer recognized the comic concern about fear of selecting the wrong beverage as one of the reasons for the experiment to be fun When all participants had completed evaluation, the beverages were decoded. Many of the respondents were astounded by what drinks they had picked as winners and what beverages they thought worst and this created a lot of amusement among the respondents.

The teacher and observers also agreed that the respondents had understood the principles factorial design experiments. However, the two observers questioned that the concept of interaction effects was fully grasped. The centre point experiment and and following discussion of curvature was probably also over most respondents level of understanding.

Results related to the Seven principles to accieve good learning:

1. High student-faculty contact:
The teacher had succeded to attain good contact within the hour and this was partially due to the topic seemed appreciated. The teacher also helped participants with calculation of effects of individuals.

2. Cooperation among students:
Results show that learning is best accomplished in groups, not individually. The course did at the beginning not leave much room for cooperation. Instead, cooperation was advised against so that results should be individual. However, the participants were instructed to and did cooperate to calculate effects after testing was complete.

3. Active learning:
This principle is about connecting theory to practical reality. The observers where unite in that this aspect was fulfilled.

4.Prompt feedback:
Pupils should have prompt feedback to their actions and the observers were unite in that this aspect was fulfilled as well.

5. Time on task:
TiME on task means that the more time and effort is put in by students, the more they will learn. As the course was limited to one hour, this aspect suffered.

6. High expectations:
Students perform well if expectations are high. There was no examination on the course, and no demands were presented.

7. A respect of diverse talents and ways of learning:
Different individuals learn n different ways. The one-hour course was streamlined to fit into the hour at disposal so this principle was also overlooked.

Survey results

The One-Hour-Course was included in a package of quality and environmentally directed courses that filled one day and at the end of the day, all participants were surveyed of their impressions of the day at the campus and were asked to rate their visit. They had to rate the

full day course on a 5 grade Likert scale and give comments on what was particularly good or particularly bad with the day and also if they had other reflections to share. On the final survey of the day, 2 of ten respondents gave the day grade five and the rest graded the day 4 on a five grade scale, where five was best. Four of the participants specifically mentioned the Cola experiment in their survey and all comments regarding the experiment were positive .

Follow-up experiment
Based on the impressions of the observers and the teacher, a follow-up experiment where it would be possible to gain deeper insights into what participants had learnt during the one-hour course. The experiment was run the same way, except for that now one replicate of the centre point experiment was included from the start. The participants of he follow-up experiment consisted of three freshman students that participated voluntarily of their experiences. Since these students already were at campus, these could be interviewed more thoroughly after the experiment without jeopardizing retention. However, these students had had their first contact with university statistics, so what these students learned could be considered an upper limit as to what non-university pupils would learn and apprehend.

The three students that had participated in the experiment where given a questionnaire where they had to manifest what they had understood and how they had experienced the course. The students had to answer what *effects* and *interactions* were*, what +, 0, -* stood for, what *statistical significance* was, and *why centre points* where used. They also had to rate the course according to how, stimulating, interesting and how high tempo the course had, as well as how much they were eager to learn more and how serious they felt the course was. Finally, they also were asked for three positive and three negative statements about the course.

Follow up experiment results related to course experience
On a scale where 1 represented very boring and 5 represented very fun, two students rated the course 5 and one 4. Two rated the course 5 for very interesting and one rated 3. Perhaps surprisingly, all students rated the course 3 with respect to how high tempo they felt the course had. The question regarding the seriousness of the course, the students rated the course 3, 4 and 5.

When asked for three positive statements regarding the course, all students thought it was fun to evaluate for the selves. Two students answered that the mix between theory and practice was positive and two students thought the course was interesting. Other single comments where: *Instructive*, *Drinking Cola* and *The theoretical lecture.*

One student had not given any answers to what was negative about the experiment. One only answered that it was negative to have to drink beverage with artificial sweetener. The last student answered that it was a pity that it was over so quickly and that it could have been deeper.

Follow up experiment results related to comprehension
The concept of *effects* appear to have been grasped by all students. All students also understood that the *replicate of one experiment* was made to be able to compare effects and to separate active effects from effects that most likely were only due to randomness. All students had some understanding of the concept of *statistical significance*. Two students appear to have understood the c*oncepts of + and – signs as presented in the design matrix* and that the *0-level experiments* represented a level in-between. The *interaction effect* was however not grasped by any of the students.

Discussion:
The One Hour Design of Experiments course is not an experiment such that a control group of participants were tutored in different ways, chiefly because there are no generally accepted ways to give a one-hour course in statistics. Since the goal was to inspire to further studies within the hour, the outcomes in terms of knowledge and test scores can naturally not be could be compared to longer courses to which campus students apply. The study of the One Hour DoE Course is a study of complex social interactions and these complex events are interpreted by two observers that, although being instructed to be as impartial as possible, probably does bias there observations in favour for the success of the course. These objections are as just here as they are in other types of action research (Stringer 1999).

However, all of the researchers interpreted the expressions of the participants as that the course was a success and that the participants appeared to have fun and learn something about science as well as about beverage tastes. The view that the test was fun was also strengthened by the comments on the surveys handed out after the full day was completed.

In light of the Seven Principles, the One-Hour DoE course comprised a very active learning environment. The experiment also had fair to high student-faculty contact in the sense that the teacher repeatedly discussed problems or issues of the experiment with the participants as well as used one of the test results as a demonstration. The experiment also encouraged cooperation among the participants when effects were calculated. Feedback of their taste results was also very quick.

The time on task was naturally short, and the participants had naturally learned more with more time. This was however a inherent limitation of the experiment. High expectations lead to good learning such that students excel. Since the purpose of this experiment was not to maximise learning and to examine the participants, the course could be said to score poorly in this practice. The one-hour experiment did not focus on different ways of learning since this was prohibited by the time restrictions.

Conclusions:
The One-hour DoE Course shows that it is possible to demonstrate the benefits of statistical thinking to students with limited knowledge in mathematics and statistics within a limited period of time. Of the seven principles of good learning, the One-Hour Experiment was ranked high on four, which, given the time restrictions is considered good. Statistical thinking and means to introduce statistical methods is also necessary in other environments. A study by Bergquist and Albing (2005) showed that the need for statistics competence in Swedish organisations is large. Fun, short and sound statistical demonstrations, followed by more extensive training could be a key ingredient in an effort to introduce statistics.

References

Andersson-Cook, C.M., (2000), "Designing a First Experiment: A Project for Design of Experiment Courses," *The American Statistician*, 52, 338-342

Bergquist, B. & Albing, M. 2005. Statistical methods – Does anyone really care, *submitted for publication*.

Bisgaard, S. 1991. Teaching Statistics to Engineers, *The American Statistician*, 45, 274-283

Chickering, A. W. & Gamson, Z. F. 1987. Seven Principles for Good Practice in Undergraduate Education*, AAHE Bulletin,* 39, no. 7:3-7

Hopkins, D. 1993. *A Teachers Guide to Classroom Research - 2nd edition*. Open University Press.

Lawrence, A. J. 1996. A Design of Experiments Workshop as an Introduction to Statistics, *The American Statistician,* 50, 156-158

Macintyre, C. 2000. *The Art of Action Research in the Classroom*, David Fulton Publishers.

Stringer, E. T. 1999. *Action Research - 2nd edition.* Sage Publications.