An Online System for Teaching the Design (and Analysis) of Experiments

C. J. Schwarz
Department of Statistics and Actuarial Science,
Simon Fraser University
Burnaby, BC, Canada V5A 1S6
cschwarz@stat.sfu.ca

July 23, 2003

Keywords: computer simulation; virtual experiments.

Abstract

Most books on experimental design have many exercises at the end of chapters that give students practise in the *analysis* of completed experiments, but students often receive little experience in the *design* of experiments. Many authors have suggested the use of student designed and run experiments, but this can be logistically costly and time consuming. Alternatively, computer generated data can be used where every student receives different response values. The difficulty in this approach lies in making simulated data easily available to students. This paper presents a virtual construction machine that, based upon information in a control file, generates a virtual laboratory that is web-accessible and easily used by students. It is very flexible and can accommodate different design objectives by varying the control parameters. A key advantage of this approach is that students can quickly try many different designs on the same experiment and learn to distinguish among designs based on randomization choices.

1 Introduction

Students learn statistics by doing statistics (Smith, 1998; Rossman and Chance, 1999). For students in introductory courses, this had lead to much emphasis on hands-on activities where students collect data through simple experiments (e.g. Jowett and Davis, 1960; Mackisack, 1994; Martinez-Dawson, 2003; Schaeffer, 1996). But what about teaching about the actual "design of experiments" in more advanced classes?

In a prescient article, Hunter (1977) stated it best, "In most courses on experimental design, students get no practice *designing* experiments, although from home work assignments, they do get practice *analyzing* data." He further describes a series of actual experiments done by students in his classes that could have complex designs.

Hunter (1977) encapsulates nicely the two extremes to student exercises in the design and analysis of experiments.

On one hand, are exercises with pre-existing, pre-run experiments. All students in a classroom are assigned the same question and all obtain the same results. These types of questions have some unfortunate, unintended consequences (Vaughan, 2003):

- students may believe that there is one, unique answer to a scientific problem.
- students do not see the extent that natural variation can change results when the identical experiment are repeated. For example, while they may compute a standard error for an estimate, they don't see that estimates do indeed vary among their fellow students.
- students may believe that results are clear-cut because everyone always gets the same answers.

On the other extreme, are actual activities where students collect and analyze their own data. These activities demonstrate to students that indeed data are variable, and that different scientists doing identical experiments may receive different answers. The primary disadvantage of these activities is logistical – materials and supplies must be available for every student. While many institutions are willing to invest in laboratories for the classical physical sciences, they are less willing to make the same investment for statistical laboratories. A secondary disadvantage is that these activities are time consuming to run and there may be insufficient time and resources for students to repeat the experiment using different experimental designs to compare and contrast both the design and analysis phases.

Hunter (1977) suggests using simulated data where "each student receives data based upon the design he or she chooses." Furthermore, he states "if the teacher uses simulated data creatively, students can learn many important aspects of design and analysis in a most effective way . . . the untapped potential is enormous . . . ." Recently, Vaughan (2003) used a similar idea where student-specific datasets are distributed via e-mail to individual students in a class for simple exercises involving sampling distributions.

Desirable features of such computer-based experiments are:

- web-based so that it is platform independent and available to students both at school and at home;
- capable of many different experimental designs;
- general enough that more just one design is "suitable" for any particular experiment;
- each replication of the experiment should return different results even if the same student uses the exact same design;
- responses should be programmable by the instructor to emphasize unique features of an experiment, e.g. large among-subject variation compared to within-subject variation; the size of the interaction among factors compared to main effects; missing data can be generated.
- easily setup by an instructor without detailed knowledge of web programming;

Such types of computer laboratories have been constructed for survey sampling. For example, Schwarz (1997) created a web-based online-village (STAT-VILLAGE) using census micro-data. Chang et al. (1992) and Lohr (1999) created a program (SURVEY) to return simulated results from a fictitious county. In both cases, students can select units according to a wide variety of sampling plans and receive back individualized data sets.

Mills (2002) reviewed using computer simulation methods to teach statistics but failed to identify any prior work similar to that proposed in this paper. A web search also failed to identify any similar efforts in experimental design, except for Anderson-Cook and Raj (2001) who designed a simulation experiment to compare a "good" and "poor" experimental design.

This paper presents an alternative to physical experiments based on computer simulation. Students access a virtual laboratory via the web where they have freedom to design their experiments. Pseudo-random numbers ensure that all students receive individualized responses, even if the exact same design was run twice in a row. The laboratories can be easily constructed by an instructor to allow for different number of levels, different blocking arrangements, and different sources of variation.

2 Student Perspective

A simple illustrative experiment will demonstrate how the interactive [1] system works and will show how a relatively simple experiment can be used to

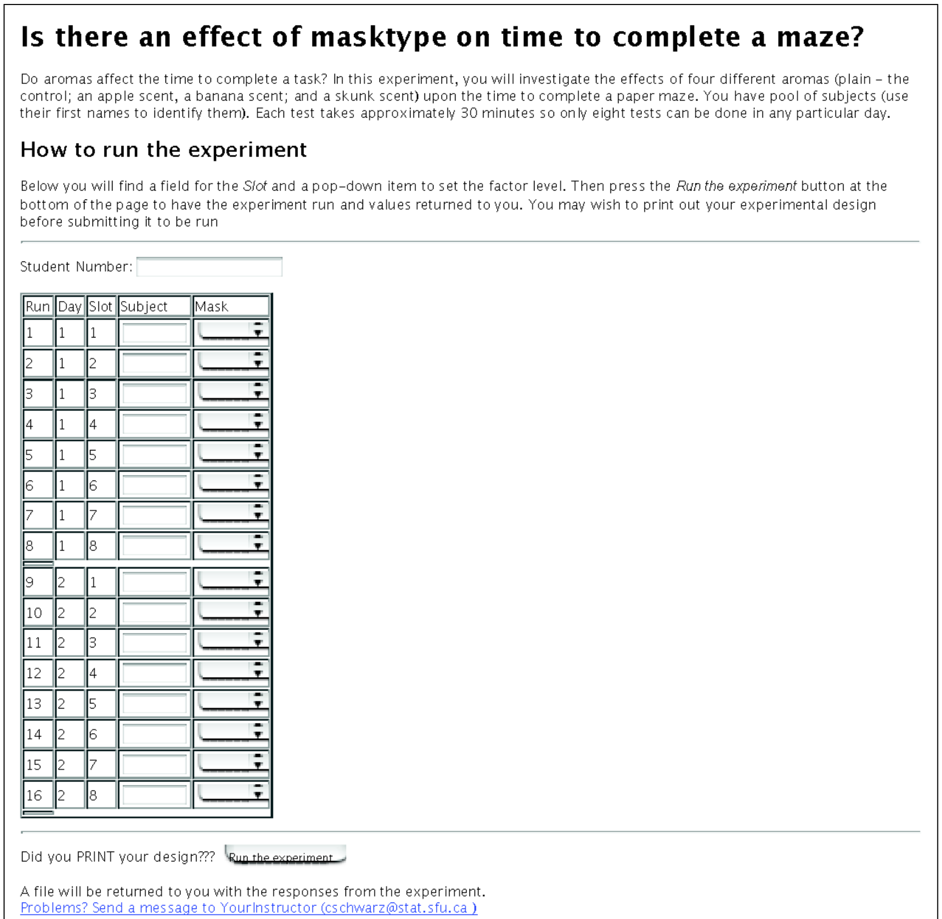
[1] Available from `http://www.stat.sfu.ca/~cschwarz/ExpDesign`

quickly demonstrate several different experimental designs for a single factor experiment.

The context is an imaginary experiment based on Hirsch and Johnston (1996). An experiment is to be conducted to investigate if different scented face masks had an effect upon the mean time to complete a maze using pen and paper. In the illustrated implementation, four levels of scent will be used: plain (unscented) as a control; two different fruit scents; and skunk scent. Because each trail of the experiment "takes approximately 1/2 hour", only eight runs can be conducted in "day". Two days are available for experimentation, i.e at most 16 runs of the experiment can be done.

Access to the basic laboratory is via a web page and the basic laboratory is shown in Figure 1. [2]

[2] The reader may try the experiment at `http://www.stat.sfu.ca/~cschwarz/ExpDesign/Scent1`



Is there an effect of masktype on time to complete a maze?

Do aromas affect the time to complete a task? In this experiment, you will investigate the effects of four different aromas (plain – the control; an apple scent, a banana scent; and a skunk scent) upon the time to complete a paper maze. You have pool of subjects (use their first names to identify them). Each test takes approximately 30 minutes so only eight tests can be done in any particular day.

How to run the experiment

Below you will find a field for the *Slot* and a pop-down item to set the factor level. Then press the *Run the experiment* button at the bottom of the page to have the experiment run and values returned to you. You may wish to print out your experimental design before submitting it to be run

Student Number:

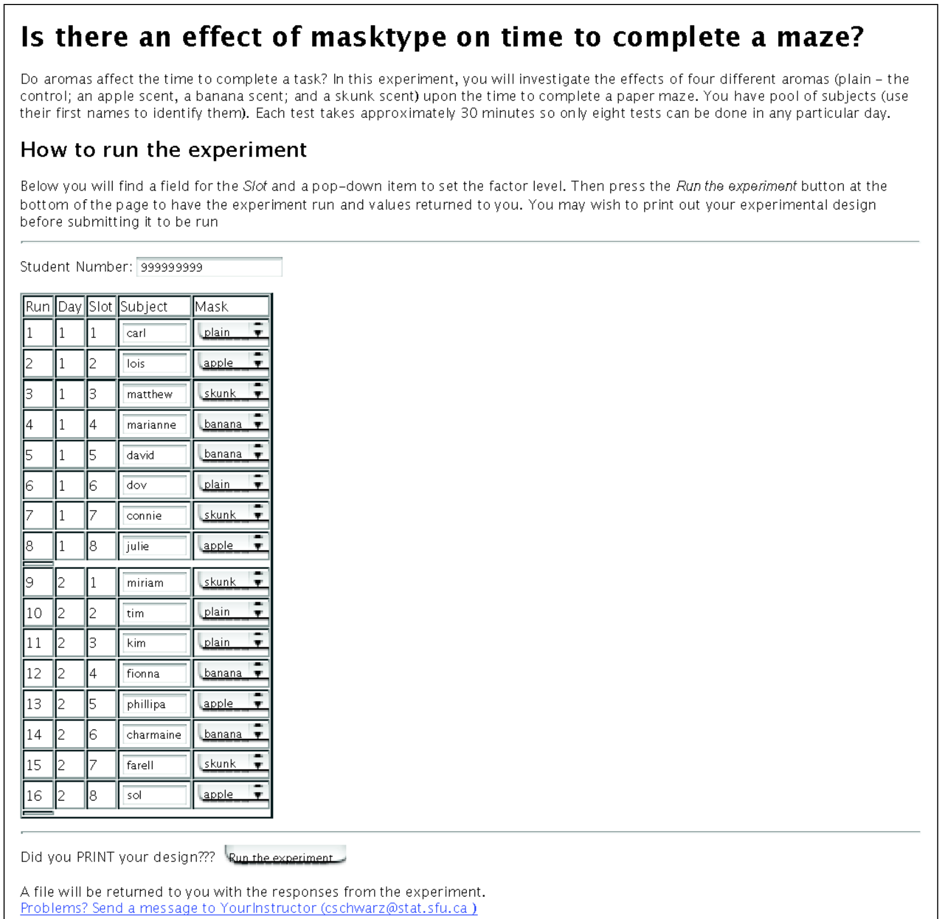
Run	Day	Slot	Subject	Mask
1	1	1		
2	1	2		
3	1	3		
4	1	4		
5	1	5		
6	1	6		
7	1	7		
8	1	8		
9	2	1		
10	2	2		
11	2	3		
12	2	4		
13	2	5		
14	2	6		
15	2	7		
16	2	8		

Did you PRINT your design??? Run the experiment

A file will be returned to you with the responses from the experiment.
Problems? Send a message to YourInstructor (cschwarz@stat.sfu.ca)

Figure 1. The virtual laboratory prior to running an experiment. Students must enter their student number, complete the text boxes to identify subjects, and select the treatments to be allocated to subjects using the pop-down menu.

To run an experiment, a student must decide upon how many subjects to use, how to randomize the treatments to the subjects, and how to randomize the subjects among the 16 runs. For example, Figure 2 illustrates how one student might perform a completely randomized design (CRD) using 16 different subjects.



Is there an effect of masktype on time to complete a maze?

Do aromas affect the time to complete a task? In this experiment, you will investigate the effects of four different aromas (plain - the control; an apple scent, a banana scent; and a skunk scent) upon the time to complete a paper maze. You have pool of subjects (use their first names to identify them). Each test takes approximately 30 minutes so only eight tests can be done in any particular day.

How to run the experiment

Below you will find a field for the *Slot* and a pop-down item to set the factor level. Then press the *Run the experiment* button at the bottom of the page to have the experiment run and values returned to you. You may wish to print out your experimental design before submitting it to be run

Student Number: 999999999

Run	Day	Slot	Subject	Mask
1	1	1	carl	plain
2	1	2	lois	apple
3	1	3	matthew	skunk
4	1	4	marianne	banana
5	1	5	david	banana
6	1	6	dov	plain
7	1	7	connie	skunk
8	1	8	julie	apple
9	2	1	miriam	skunk
10	2	2	tim	plain
11	2	3	kim	plain
12	2	4	fionna	banana
13	2	5	phillipa	apple
14	2	6	charmaine	banana
15	2	7	farell	skunk
16	2	8	sol	apple

Did you PRINT your design??? Run the experiment

A file will be returned to you with the responses from the experiment.
Problems? Send a message to YourInstructor (cschwarz@stat.sfu.ca)

Figure 2. A possible experimental design. This is a single factor completely randomized design (CRD). The student then clicks on the "Run" button to get the experiment data shown in Figure 3.

Notice that the student must decide upon the randomization protocol at all levels, i.e. assigning treatments to subject and the run-order. Once the design is complete, the student clicks upon the *Run* button at the bottom of the page and the “data” from the experiment are returned (Figure 3)

Here are the results of your experiment.

Save this file as a text file using the File menu and the Save options.
Then strip off the lines before and after the raw data.
Be sure that the file is saved as a TEXT file.

```
Student number, Day , Slot , Subject , Mask , Response (Y)
999999999 01 01 carl plain 201
999999999 01 02 lois apple 183
999999999 01 03 matthew skunk 198
999999999 01 04 marianne banana 155
999999999 01 05 david banana 184
999999999 01 06 dov plain 163
999999999 01 07 connie skunk 173
999999999 01 08 julie apple 161
999999999 02 01 miriam skunk .
999999999 02 02 tim plain 107
999999999 02 03 kim plain 129
999999999 02 04 fionna banana 127
999999999 02 05 phillipa apple 140
999999999 02 06 charmain banana 125
999999999 02 07 farell skunk 110
999999999 02 08 sol apple 157
End of results listing
```

Figure 3. The results from the experiment run in Figure 2. If the same experimental design were to be run, different values would be returned. This file should be saved to a text file for further analysis by the student.

The student will save this information to a text file for subsequent analysis. If the exact same experimental plan as "rerun", different responses will be returned.

Notice that this simple laboratory allows for several types of experimental protocols:

- a completely randomized design (CRD) as illustrated in Figure 2.
- a simple randomized complete block design (RCB) using morning and afternoons as blocks. Students must ensure that every treatment occurs in every block.
- a generalized randomized complete block (GRCB) design using complete days as blocks (Addelman, 1969; Gates, 1995). Now some treatments occur multiple times within a block. The advantage of GRCBs is that

they allow testing of the assumption of no block-treatment interaction (i.e. are blocks and treatments additive?).

- a sub-sampling design where the same subject-mask combination is measured more than once.
- a repeated measures design where every subject is measured using all masks.

3 Instructor's Viewpoint

The construction of such a virtual laboratory can be tedious if approached as a one-off task. It is also difficult to do unless a person has good web-programming skills. However, a "virtual construction company" is available that can construct a virtual laboratory for a wide range of alternative plans based upon a control file. The control file used to build the laboratory illustrated in Figure 1 is shown in Figure 4. [3]

[3] The control file is available at `http://www.stat.sfu.ca/~cschwarz/ExpDesign/Create`

```
Line  Control Field
1     Is there an effect of masktype on time to complete a maze?
2     Do aromas affect the time to complete a task? In this experiment, you will
3     investigate the effects of four different aromas (plain - the control; an apple
4     scent, a banana scent; and a skunk scent) upon the time to complete a paper maze.
5     You have pool of subjects (use their first names to identify them). Each test takes
6     approximately 30 minutes so only eight tests can be done in any particular day.
7     END of instructions - important that first 3 letters say END in uppercase.
8     1                         - number of factors
9     4 Mask                    - number of levels for factor and name of the factor
10    plain    140   0            Then for each level, enter its label, the average,
11    apple    145   0            and the increase in the standard deviation
12    banana   140   0
13    skunk    155   2
14    2 8                      - number of blocks and  block size
15    Day                      - name of "blocks"
16    15                         Std deviation of block effect
17    Slot                     - name of units within blocks
18    5                          Std deviation of responses at unit level (MSE)
19    Subject                  - name of experimental unit
20    20                         Experimental unit variation (subject effects)
21    .10                      - fraction of missing values
22    InstructorName           - Instructor name
23    cschwarz@stat.sfu.ca     - instructor email address
24    no                       - trace flag (used for debugging) yes or no
25    yes                      - email flag (send student design to instructor via email?)
```

Figure 4. Control file used to generate the virtual laboratory in Figure 1.

Line 1: Title of the experiment
Lines 2-7: Explanatory text that appears on web page with designed experiment
Line 8: Number of factors
Line 9: The number of levels in the single factor and the factor name.
Lines 10-13. For each level, the level name, the level mean, and the increase in variation for that level.
Line 14: The number of blocks and block size.
Line 15: The name of the blocking variable.
Line 16: The standard deviation of the blocking effect.
Line 17: The name of the runs.
Line 18: The standard deviation of the run effect.
Line 19: The name of the experimental unit.
Line 20: The experimental unit standard deivation.
Line 21: The fraction of missing values
Line 22: Instructor name and email address.
Line 23 The instructor's email address
Line 24: A trace flag used for debugging purposes – ordinarily not used.
Line 25: Does the instructor want the student design and responses emailed to them?

parameters. First, there are those that control the physical layout, e.g. the factor name, the number of levels, the level names, how many blocks, the name of the blocking factor, etc. Second, there are parameters that control the generation of responses. The underlying statistical model is:

$$Y_{ijkl} = \mu_i + b_j + s_k + r_l + v_i$$

where

- Y_{ijkl} is the response of treatment level i, when run in block j, subject k in run location l. Not all combinations of i, j, k, l may be present in an experiment;
- μ_i is the mean response for treatment level i;
- b_j is the effect of block j assumed to be normally distributed with a mean of 0 and a variance σ^2_{Blocks};

- s_k is the effect of subject k assumed to be normally distributed with a mean of 0 and a variance of $\sigma^2_{subject}$;
- r_l is the effect of run-order position l assumed to be normally distributed with a mean of 0 and a variance of $\sigma^2_{runorder}$;
- v_i is the excess variation associated with treatment level i assumed to be normally distributed with a mean of 0 and a variance of σ^2_{excess}.

Missing data occurs completely at random (MCAR) at a rate controlled by the instructor.

By a judicious choice of the parameters, it is easy to design a virtual laboratory with different properties. For example,

- $\sigma^2_{Blocks} = 0$ implies no block effects; $\sigma^2_{Blocks} >> \sigma^2_{subject} + \sigma^2_{runorder}$ implies a large block effect and so a block design would be more efficient.
- $\sigma^2_{subject} >> \sigma^2_{runorder}$ implies a large among subject variance relative to within-subject variance where a paired/split-plot design would be preferable.
- if all μ's are comparable and smaller than the variance terms, the design will have low power.
- $\sigma^2_{excess} \neq 0$ implies data where the assumption of homogeneous variances is violated.
- the ratio of $\sigma^2_{subject}$ to $\sigma^2_{runorder}$ controls if sub-sampling (multiple runs on the same subject with the same treatment)is desirable;
- in multifactor designs, the treatment means can be set to obtain any desired size of interaction among the factors.

4 Underlying mechanics

The following is a overview of the mechanics of creating and what happens when a student uses the virtual laboratory. Detailed instructions that come with the program files should be consulted for actual installation details.

4.1 Construction phase

1. The instructor completes the control file using a simple text editor.

2. The instructor sends the control file to a the "virtual construction company" from a web page. [4]

3. The "construction company" is a a *SAS* program that reads the control file and creates an *HTML* files that produces the web page as illustrated in Figure 1. This web page contains hidden fields with information that is passed to the simulation program.

4. (optional) The instructor places a generic *PERL* script and a second *SAS* program into the Common-Gateway-Interface (CGI) section of their web-directory. This will entail a slight alteration to the HTML file to redirect the processing; otherwise the scripts on the author's institutions machines are used.

5. The instructor makes the file created by the construction machine available on the web to students.

4.2 Running experiments

1. Students complete the experimental protocol (e.g. Figure 2) and click on the *Run* button.

2. The data on the form and the hidden control fields are passed to the *PERL* script. This script parses the information, generates responses according to the statistical model, and passes the responses back to the student.

The *PERL* scripts and *SAS* program are general enough to handle any virtual laboratory created in the construction phase and need not be modified by the instructor.

5 More complex designs

While a single factor designs are sufficient to illustrate the concepts of blocking, subsampling, and repeated measures, they are unable to illustrate perhaps the most complex design structure - the split-plot design. Consequently, a similar virtual laboratory can also be constructed using the same basic principle for two factor designs. This is illustrated for the mask experiment in Figure 5. [5]

[4] `http://www.stat.sfu.ca/~cschwarz/ExpDesign/Create`

[5] The page is available at `http://www.stat.sfu.ca/~cschwarz/ExpDesign/Scent2`

Is there an effect of masktype on time to complete a maze - II?

Do aromas affect the time to complete a task? In this experiment, you will investigate the effects of four different aromas (plain - the control; an apple scent, a banana scent; and a skunk scent) and gender (male or female) upon the time to complete a paper maze. You have pool of subjects (use their first names to identify them). Each test takes approximately 30 minutes so only eight tests can be done in any particular day.

How to run the experiment

Below you will find a field for the *Slot* and a pop-down item to set the factor level. Then press the *Run the experiment* button at the bottom of the page to have the experiment run and values returned to you. You may wish to print out your experimental design before submitting it to be run

Student Number: 999999999

Run	Day	Slot	Subject	Gender	Mask
1	1	1	carl	male	plain
2	1	2	lois	female	apple
3	1	3	matthew	male	plain
4	1	4	marianne	female	skunk
5	1	5	david	male	skunk
6	1	6	samantha	female	apple
7	1	7	meghan	female	skunk
8	1	8	paul	male	apple
9	2	1	joan	female	banana
10	2	2	mark	male	banana
11	2	3	shawna	female	plain
12	2	4	stephen	male	skunk
13	2	5	carolyn	female	plain
14	2	6	lynn	female	banana
15	2	7	richard	male	plain
16	2	8	stephen	male	banana

Did you PRINT your design??? Run the experiment

A file will be returned to you with the responses from the experiment.
Problems? Send a message to YourInstructor (cschwarz@stat.sfu.ca)

Figure 5. A possible experimental design for a two factor experiment. This is a two factor completely randomized design (CRD). The student clicks on the "Run" button to get the experimental data.

This virtual laboratory can be used to run CRD, RCB, GRCB and variants as before. A split-plot designs can be run by holding the subject fixed within a sex, and running all mask-types on this subject in random order.

Generation of these virtual laboratories is again done using a control file and is not illustrated in this paper. [6]

[6] A copy of the control file is available at `http://www.stat.sfu.ca/~cschwarz/ExpDesign/Create`

6 Student experiences

I have used virtual laboratories for assignments and take-home final exams for a course in experimental design for environmental science students for the last few years [7] The exams are supposed to be as realistic as possible - students must first collect some preliminary data (i.e. a pre-test), use this preliminary data to decide upon appropriate sample sizes and designs, collect data from their designed experiment, and write a report in the form of a statistical paper. This makes the planning and execution as similar as possible to the the steps required with real experiments as outlined in Short and Pigeon (1998). In the past, generation of these take home exams was done on a one-off basis, but this became tiresome, and inspired this paper.

While no formal evaluation of this method of testing students has been performed, my anecdotal evidence points to surprise among the students that people that run identical designs can come with quite "different" point estimates and p-values. Indeed, my experience is that some students (about 5% of them) are surprised to find evidence of differences when nearly everyone else in the class fails to detect differences in the mean response! As delMas et al. (1999) note, the development of anecdotal evidence is the first step in a longer term "classroom research" project.

7 Discussion

There are several obvious advantages of the virtual laboratory over real physical experimentation – reduced space and operating costs and the ability to run several different designs on the same experiment. However, the most important advantage of the virtual laboratory over real physical experimentation is ease of control. Designing a real physical experiment is hard and, in many cases, must be done using trial and error to ensure that effects are not obvious and that it is robust to student usage. For example, while it is not difficult to construct a similar physical experiment to the ones presented in this paper, it is difficult to know the size of the person-to-person variation, and difficult to ensure that every student has sufficient subjects. How big are blocking effects? How can missingness be controlled? All of these are easy to do by changing values in the control file.

Another advantage is that is relatively easy to repeat an experiment using different experimental protocols in a short period of time. For example, it is quite instructive for student to compare the randomization process between

[7] Visit `http://www.stat.sfu.ca/~cschwarz/Stat-650/Exam/study` for copies of these past exams. The TakeHome finals of recent years are of most interest

CRD, RCB, GRCB, sub-sampling, and repeated-measure designs. Furthermore, it is also quite instructive to show students that the raw data themselves often do not provide sufficient information to identify the design and that a description of the experimental protocol must always be examined.

A potential disadvantage of this approach is the "lack of realism" - do students learn from computer simulations what we want them to learn? In a recent literature review, Mills (2002) indicates that computer simulation builds upon the theory of constructivism in helping students to learn by requiring them to actively interact with the subject matter. Mills (2002) further concludes that there is little empirical support for many of the purported advantages of computer simulation models. Rossman and Chance (1999) recommend that students first start with a physical simulation before using computer based simulations as a way to connect the numbers and displays being produced with the process being simulated. In our virtual laboratory, students must still do the randomization themselves - only the physical responses are simulated - and it is precisely the different randomization procedures that is at the core of experimental *design* as opposed to the *analysis* of experimental data.

8 Acknowledgments

Research support by the National Science and Engineering Research Council (NSERC) of Canada.

9 Reference

Addelman, S. (1969). The generalized randomized block design. American Statistician, 23, 35-36.

Anderson-Cook, C.M. and Dorai-Raj, S. (2001). An Active Learning In-Class Demonstration of Good Experimental Design. Journal of Statistical Education [Online], 9(1).

Chang, T. C., Lohr, S. L., and McLaren, C. G. (1992). Teaching Survey Sampling Using Simulation. American Statistician, 46, 232-237.

delMas, R. C., Garfield, J., and Chance, B. L. (1999). A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning Journal of Statistics Education [Online], 7(3).

Gates, C. E. (1995). What really is experimental error in block designs? American Statistician, 49, 362-363.

Hirsch A.R and Johnston L.H. (1996) Odors and learning. Journal of Neurological and Orthopedic Medicine and Surgery, 17, 119-126.

Hunter, W. G. (1977). Some Ideas About Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students. American Statistician, 31, 12-17.

Jowett, G. H., and Davies, H. M. (1960). Practical experimentation as a teaching method in statistics. Journal of the Royal Statistial Society, Series A, 123, 10-36.

Lohr, S. L. (1999). *Sampling: Design and Analysis* Brooks/Cole, Pacific Grove, CA.

Mackisack, M. (1994). What Is the Use of Experiments Conducted by Statistics Students? Journal of Statistical Education [Online], 2(1).

Martinez-Dawson, R. (2003). Incorporating Laboratory Experiments in an Introductory Statistics Course. Journal of Statistical Education [Online], 11(1).

Mills, J.D. (2002). Using Computer Simulation Methods to Teach Statistics: A Review of the Literature. Journal of Statistical Education [Online], 10(1).

Rossman, A. J., and Chance, B. L. (1999). Teaching the Reasoning of Statistical Inference: A 'Top Ten' List. College Mathematics Journal, 30, 297-305.

Schaeffer, R. L. , Cnanadesikan, M., Watkins, A. and Witmer, W. J. (1996). *Activity based statistics.* Springer-Verlag, New York.

Schwarz, C. J. (1997). StatVillage: An On-Line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling. Journal of Statistical Education [Online] 5(2).

Short, T. H. and Pigeon, J. G. (1998). Protocols and Pilot Studies: Taking Data Collection Projects Seriously. Journal of Statistical Education [Online], 6(1).

Smith, G. (1998). Learning Statistics By Doing Statistics. Journal of Statistical Education [Online], 6(3).

Spurrier, J. D. (2000). *The Practice of Statistics: Putting the Pieces Together* Duxbury, Pacific Grove, CA.

Vaughan, T. S. (2003). Teaching statistical concepts with student-specific datasets. Journal of Statistical Education [Online], 11(1).