Practical Aspects of Engineering Experimentation

by

Virgil L. Anderson
Department of Statistics

and

William D. Kovacs
School of Civil Engineering

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #408

May 1975

Practical Aspects of Engineering Experimentation

by

Virgil L. Anderson
Department of Statistics

and

William D. Kovacs
School of Civil Engineering

Purdue University
West Lafayette, Indiana 47907

PRACTICAL ASPECTS OF ENGINEERING EXPERIMENTATION

by

Virgil L. Anderson and William D. Kovacs
Purdue University
West Lafayette, Indiana

Up-to-date engineering experimentation has a practical and realistic approach in which engineers choose their inference space and restrictions on randomization so that they obtain the best designed experiment.

Abstract

Engineers have used "seat-of-the-pants" methods to attack problems for many years. In the past ten years they have used a non-design of experiments method that has left something to be desired. In the 1970's they have been turning to a realistic approach in designing experiments that has been quite successful in solving complex problems as well as the usual problems.

Assuming the engineer knows what his problem is and what it is not, the realistic approach to experimentation allows the engineer to describe the inferences he wants to make from his experiment before he takes the first observation and to choose the restrictions on randomization that give him the least trouble in running the experiment for the information he must obtain. These two choices (inference space and restrictions on randomization) along with the necessary replication provide the basis for setting up the corresponding mathematical model before the experiment is run. If the model and the accompanying outline of the analysis of the data to be taken show that the engineer cannot obtain the information he wants, he can change the design (possibly the inferences and/or restrictions on randomization) in such a manner that he is assured of the correct attack on the problem.

Present practices allow almost no one-factor-at-a-time experiments because no interaction information is obtained. More recently, many factor experiments have been used rather than many experiments with a couple of factors in each. The reasons include not only the information on the two factor interactions but also the information on the three factor interactions (as well as that on the main effects).

PRACTICAL ASPECTS OF ENGINEERING EXPERIMENTATION

by

Virgil L. Anderson
and
William D. Kovacs

Introduction

In order to solve many engineering problems in an efficient manner, the modern engineer turns to the more up-to-date experimental techniques imbedded in the scientific method. The most improved of these techniques appears to revolve around the realistic approach to designing experiments. A non-design of experiments method to attack problems has been put forth within the last ten years dealing with asking questions of when, where, how much and what changes occurred to create the problem. This method is more efficient than the so-called "seat-of-the-pants" techniques, but does not give a satisfactory solution to many engineering problems. The design of experiments method utilizes the more recent ideas and adds another dimension in a more comprehensive attack. This approach is outlined by McLean, et al (1), (Numbers in parentheses pertain to references at end of paper) while more modern designs are given by Anderson and McLean (2).

Non Design of Experiments Method

Many organized and many not so organized methods of attacking problems have been used by engineers. One of the more organized procedures used in the last ten years was suggested by Kepner and Tregoe (3). The methodology is non design of experiments oriented but suggests that

a) A problem is a deviation of what should be and what actually is.

b) A change of some kind is always the cause of the problem.

c) To find the cause one must ask certain types of questions regarding observable changes such as (where and where not, what occurred and what did not occur and how great a change occurred?)

d) The question, "Why?" is postponed or never asked to prevent too many immediate answers that may mislead the investigation.

The possible actions taken by this method are:

a) Interim, do something positive before the cause is found.

b) Adaptive, the cause is found but it is too expensive or time consuming to correct it at that time, or

c) Corrective, the best solution because the cause is found and corrected.

Design of Experiments Method

While the above procedure is not unreasonable for handling some of the problems encountered by many engineers, there is a more reliable and efficient procedure to attack their problems. The one method currently used by more progressive engineers is the realistic approach to designing experiments. This method incorporates a mathematical basis plus the practical aspects of attacking problems by using experimentation and is called the design of experiments method.

This design of experiments approach is imbedded in the overall scientific method of experimentation as follows:

a) Recognition that a problem exists.

b) Formulation of the problem.

c) Agreeing on factors and levels to be used in the experiment.

d) Specifying the variables to be measured.

*e) Definition of the inference space for the problem.

*f) Random selection of the experimental units.

*g) Assignment of the treatments to the experimental units.

h) Outline of the analysis corresponding to the design before the data are taken.

i) Collection of the data.

j) Analysis of the data.

k) Conclusions and/or interpretation.

l) Implementation.

where * (e, f and g) indicates the location of the actual design of the experiment imbedded in the scientific investigation. The parts a) through d) prepare for the design and h) allows the experimenter to modify his design before collecting the data. Next, one may consider the practical aspects of engineering experimentation by investigating the three topics e), f) and g) in detail.

Inference space is a conceptual space. The investigator carefully outlines the extent of this space to which the results of the coming experiment will apply. For example, in an experiment with soil borings using impact penetration methods, an engineer wants to compare the effects of safety drivers vs. cat head drivers, various hammers, fall height, number of rope turns on cat head and other factors on the impact velocity of the falling weight. The kinetic energy of the falling weight may then be computed from the measured impact velocity. The results would be applicable to this set of conditions with the understanding that if major effects showed up, they may be further investigated in some soil drilling operations. The inference space encompasses the variable of this experiment plus an implication of a direction to follow in field operations later. This narrowness of the actual inference space from the results of this experiment had to be recognized before performing laboratory experiments and taking the data. The reason being that if the engineer believed the results would be too restricted for a drilling contractor, he must change the study before it was begun. All too often investigators do not explicitly outline their inference space before experiments are designed and do not know the extent to which their results apply after the results are obtained.

Conclusion: The inference space must be known by the experimenter before he takes the first measurement.

Randomization allows the experimenter to make probability statements about his results. The randomization methodology should allow all levels of all factors an equal chance of showing their value in causing the response. For example, in the impact penetration experiments, the engineer wanted the safety driver to have as good a chance to show how much impact velocity it could cause as the cat head driver; likewise to give the various fall heights an equal chance to show how much impact velocity each could cause and so on, plus all combinations of drivers with fall heights and all other factor levels. The concept is all inclusive in the sense that any combination of the levels of all the factors has an equal chance to show what it can cause on the dependent variable (e.g. the impact velocity).

To accomplish this requires careful thought not to unconsciously favor one combination over another. For example, let us say that it is easier to set up the short fall height before the long fall height combinations in this impact penetration experiment. If the experimenter does this and the operator is fresher in the morning when the experiment starts, he may get higher velocity with the short fall heights due to the operator's physical condition and not because of the fall height from which an erroneous conclusion could be drawn. There are many such peculiarities in this and other experiments that may give preference to some combinations run early, late or in some other manner. Hence, it is necessary to "scramble" the combinations to eliminate the preference of a run as much as possible. Randomization is the solution here. It must be followed mechanically with tables to dictate a procedure not to allow a person to make the choice as to which combination should be run first, second and so on, for all treatment combinations.

Conclusion: In all experiments from which probability statements are to be made there must be randomization to allow valid estimates and/or tests of hypotheses.

Replication is required in experiments where no previous estimate of the variation associated with the units (from which the information is obtained) is available. This variation (called "error") from experimental units treated alike, forms the basis for the tests of significance and the confidence intervals on the parameters being investigated. For example, in the impact penetration study, the experimenter repeated each combination of levels of the factors four times. This allowed him to compare the arithmetic means of the impact velocities from the various levels of the factors with one another. The comparison is based on how large the variation of the impact velocity was from the repetitions of all the runs for each combination of levels of the factors.

When enough information is available from previous studies to provide the information on this "error" variation within combinations of levels, there is no need to replicate. In fact, if further information is available on the effect of combinations of factors, called interaction, such that the higher factor interactions (where more than two factors act together) can be assumed zero, there can be fewer observations or experimental units than the total number of combinations of levels. This condition in an experiment is called "fractional replication".

Conclusion: Of the three ingredients necessary for conducting an experiment (Inference Space, Randomization and Replication), replication is the least important for engineers because so often knowledge of the inherent variation of the experimental units is known or easily estimated. If however, no such information is available, then there must be replication in the experiment to provide an estimate of the "error", the basis for interval estimates (confidence intervals) and tests of hypotheses.

Example

In recent experiments on impact penetration in soils one may demonstrate the design of experiments principles all the way through to the analysis of some of the data and conclusions from it. Details of the experiments are described elsewhere (5).

At the design of experiments stage of the investigation, the inference space was assumed to be such that the results would apply to similar types of drilling equipment used in this study with an implication of what soil boring contractors may expect with impact velocity in the field when using this or similar pices of equipment and these levels of factors.

In order to allow the inferences to be valid, the randomization had to be complete. For example, when using the Safe-T-Driver on automatic mode and on manual mode were to be compared for zero inclination of the drill stem from the vertical, and each of these modes were to be used on a 27", 30" and 32" fall, a total of six tests are needed. For simplicity, consider these six combinations of mode and fall and how one would allow one replication of each combination in the experiment. In total, there are 12 combinations to run and are given in Table 1:

TABLE 1

Sequence of Tests

test	combination		order of performing experiment
	mode	fall	
1	auto	27"	5
2	auto	30"	4
3	auto	32"	7
4	manual	27"	10
5	manual	30"	12
6	manual	32"	2
7	auto	27"	9
8	auto	30"	8
9	auto	32"	1
10	manual	27"	3
11	manual	30"	11
12	manual	32"	6

To randomize the testing order, the engineer uses a table of random numbers and assigns them to the test number. For example, test 9 received the random number 1; then, test 9, automatic mode at 32" fall would be run first. If the next random number turned out to be 6, then manual mode at 32" would be run second and so on until the last random number assigned to test is 5 which means that the manual mode at 30" fall would be run. If the experiment had actually been run this way the model appropriate for analyzing the data would be:

$$\begin{array}{ccccccccccc} 12 & = & 1 & + & 1 & + & 2 & + & 2 & + & 6 \\ y_{ijk} & = & \mu & + & \underset{\uparrow}{M_i} & + & \underset{\uparrow}{F_j} & + & \underset{\uparrow}{MF_{ij}} & + & \varepsilon_{(ij)k} \end{array} \qquad \begin{array}{l} i = 1,2 \\ j = 1,2,3 \\ k = 1,2 \end{array} \qquad \text{Model 1}$$

where:

Numbers above the model = degrees of freedom, indicate the apportioned breakdown of the information from the 12 observations,

y_{ijk} = the impact velocity for the k^{th} run of the j^{th} fall height and i^{th} mode,

μ = overall mean fall impact velocity,

M_i = the effect of the i^{th} mode,

F_j = the effect of the j^{th} fall height,

MF_{ij} = the effect of the interaction of the i^{th} mode and j^{th} fall height,

$\varepsilon_{(ij)k}$ = the random error associated with the k^{th} run of the ij^{th} combination of mode and fall height.

The analysis of the data then uses the variation associated with ε to compare with the variation of M, F, and MF (indicated by the dashed lines below the model). If the variation for M is "significantly" larger than that for ε, the engineer then has a basis for selecting the mode that caused the larger average impact velocity. Similar analyses can be run on F and MF. Further discussion on this example (with numbers) is provided later. Here, however, emphasis is given to the randomization procedures and their effects on the models and the corresponding analyses.

Usually engineers do not want to run completely randomized experiments because they are too time consuming and expensive. Assume that there was not complete randomization in this experiment because it required too much time to disassemble and assemble the rigging for each test condition. Only the six combinations of mode-fall were randomized and once a combination, say auto mode-32" fall was drawn, both runs were made, one immediately after the other for the combination. The rigging was then dissembled and the second randomly drawn combination, say manual-27", was set up and both runs made before the third combination run. Continue in this manner until the six combinations with two runs per combination were made. This approach gives the same number of data points as shown in Table 1. A reasonable model appropriate for analyzing the data from this "restricted" randomized design is:

$$\begin{array}{lllll} 12 = 1 + 1 + 2 + 2 + 0 + 6 & & i = 1,2 & & \\ y_{ijk} = \mu + M_i + F_j + MF_{ij} + \delta_{(ij)} + \varepsilon_{(ij)k} & & j = 1,2,3 & & \text{Model 2} \\ & & n = 1,2 & & \end{array}$$

where the symbols through MF_{ij} are the same as in Model 1 but $\varepsilon_{(ij)k}$ is no longer appropriate as the basis of comparison with M, F and MF. However, the dashed line from ε stops at δ because 1.) the variation in this experiment caused by run to run differences may be too small due to constraining the randomization to the six combinations and 2.) running both repeats of one combination before running any other combination. The $\delta_{(ij)}$ is called a "restriction error" (4,2). This term indicates that the restriction of not allowing other combinations to be run until all of the runs of the first randomly chosen combination are run, and so on, is recognized in the model. Actually, $\delta_{(ij)}$ is the term that prevents a wrong test to be made on M, F and MF using the ε in this experiment. The dashed lines show that the basis for the tests of M, F and MF is δ, which has zero degrees of freedom so no tests of significance are possible as a result of this design.

The engineer is now in a dilemma. He cannot run all the combinations in a manner that allows him to analyze the data appropriately as shown in Model 1 because he wants to run the experiment in a manner that allows him to use only Model 2. How may he overcome this dilemma practically? The solution is that he must show that the variation within the combination (ε) for the experimental procedure to obtain Model 1, is as small as the variation within the combination for that to obtain Model 2. In many well controlled experiments this is true, but it must be shown in every case before it can be used for appropriate inferences. To investigate this possibility, preliminary experimentation is usually set up that allows at least one replication of the first procedure to get Model 1 with repeated runs within each combination to augment Model 1 and obtain terms as in Model 2. For example, let us set up the six combinations and run two replications as we did to obtain Model 1. Then as each of these twelve combinations comes up, run two repeats before the next combination is set up. The model to analyze the data from this experiment may be given as:

$$\begin{array}{ccccccccccccccc} 24 & = & 1 & + & 1 & + & 2 & + & 2 & + & 6 & + & 0 & + & 12 \\ y_{ijk\ell} & = & \mu & + & M_i & + & F_j & + & MF_{ij} & + & \varepsilon_{(ij)k} & + & \delta_{(ijk)} & + & \eta_{(ijk)\ell} \end{array}$$

$$\underbrace{\varepsilon_{(ij)k} + \delta_{(ijk)} + \eta_{(ijk)\ell}}_{}$$

$i = 1,2$
$j = 1\ 2,3$
$k = 1,2$
$\ell = 1,2$

Model 3

where: $\varepsilon_{(ij)k}$ is same as in Model 1
$\delta_{(ijk)}$ is similar to $\delta_{(ij)}$ in Model 2
$\eta_{(ijk)\ell}$ is as ε in Model 2,
and all other terms are same as in the other two models.

If a test of ε using η(dashed line) turns out to be very non-significant (use a probability of rejecting the hypothesis when it should be accepted at .25 which results in a small probability of accepting a false hypothesis), the engineer can then use the Model 2 type design of experiment for his whole experiment if he is willing to assume that these results will apply for all factors. In general, this is not a unreasonable assumption if the results using Model 3 truly show that variation due to ε is practically as small as the variation due to η .

Another concept dealing with the analysis of the data via the models must be understood. The concept is that much more information may be obtained on factors causing a certain response. This is true when all factors are placed in the experiment simultaneously (as in this experiment) rather than running only two factors at a time. Most engineers almost never run one-factor-at-a-time experiments anymore because all possible two factor experiments from five factors would require ten experiments. In addition to the information on the two factor interactions (which are usually the most informative) the engineer can obtain estimates of the effects of three, four and five factor interactions if the overall five factor experiment is run.

As indicated previously, the experiment on impact velocity did include many factors simultaneously; but to examine a few of the results after finding out that the term MF_{ij} in the experiment was significant, Table 2 shows average (arithmetic mean) impact velocity caused by the two factors, height of the fall and mode of using the Safe-T-Driver and Safety Hammer combination. The table is composed of a total of 111 data points from Series I & II (5).

TABLE 2

Impact Velocity (in./sec.): Fall by Mode*

	FALL		
Mode	27"	30"	32"
Automatic	115	119	124
Manual	117	124	127

*Nearest whole number

The data in Table 2 are shown in Figure 1, Impact Velocity vs. Fall. Although, the relationships plot adjacent to each other (parallelism) for the combined data, non parallelism of the two curves occurs when Series I <u>or</u> Series II is plotted separately over fall for automatic and manual driving modes. This indicates interaction of fall by modes MF_{ij} in the experiment, however it seems sufficient to just describe a couple of the inferences the engineer may draw from these data:

1) As expected, the impact velocity increases as the fall height increases.
2) The use of automatic mode over manual mode results in a lower impact velocity for the range of fall studied. (Later tests in which the

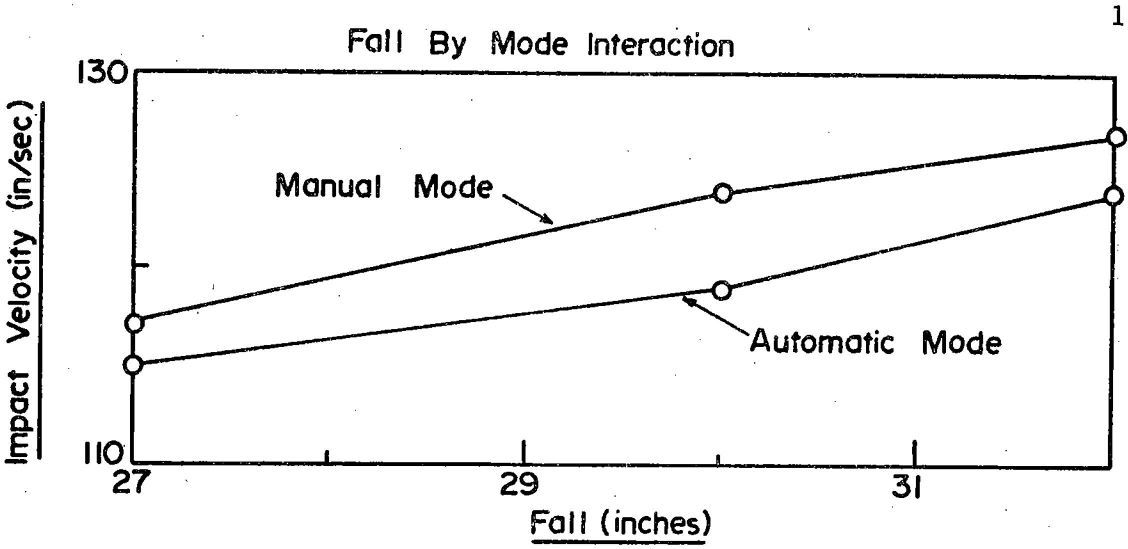

FIGURE I IMPACT VELOCITY VS. FALL FOR SAFE-T-DRIVER AND SAFETY HAMMER FROM SERIES I & II.

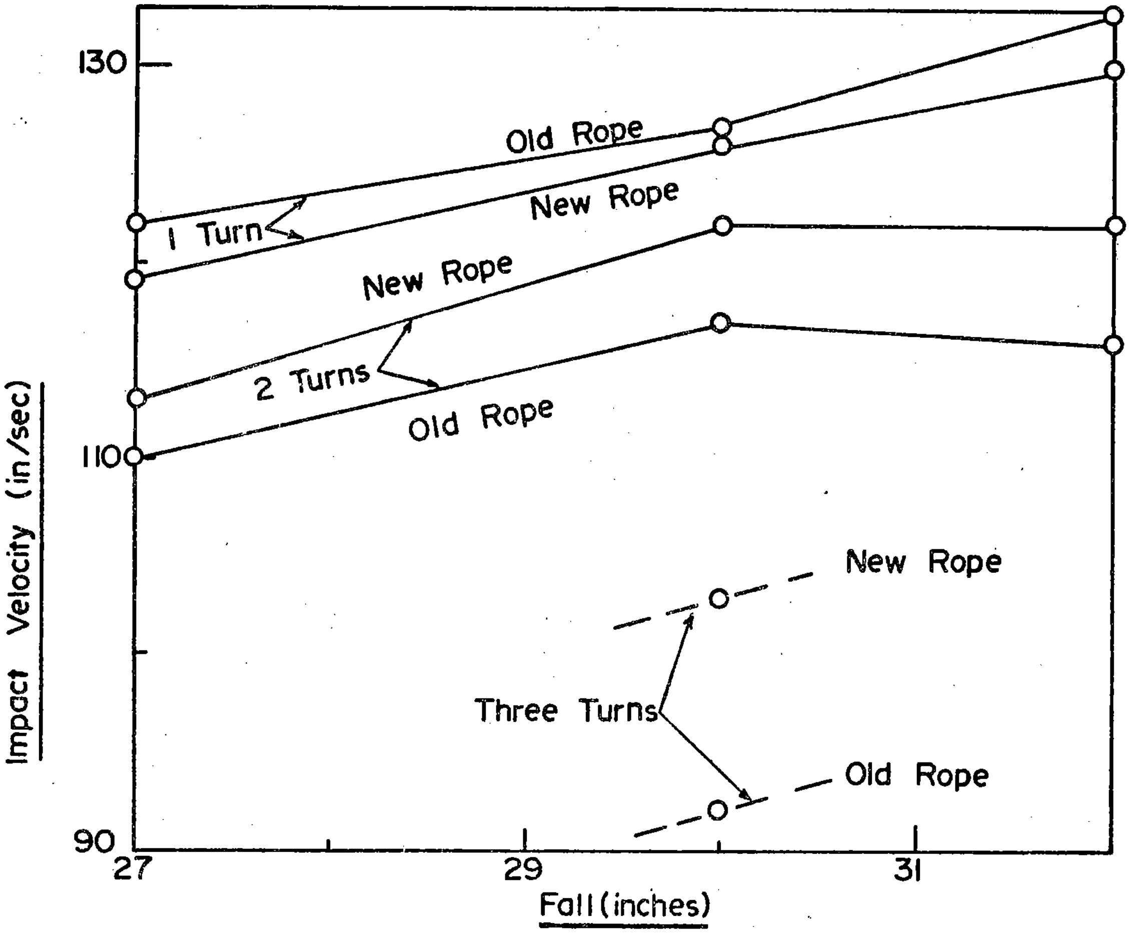

FIGURE 2 IMPACT VELOCITY VS. FALL FOR CATHEAD WITH ROPE AND CYLINDER WEIGHT–THREE FACTOR INTERACTION.

slack was adjusted for the automatic mode showed equal impact velocity between modes.

3) For a 30" fall (which is the prescribed fall for performing the impact penetration test) the automatic mode is only about 4% less than the manual mode.

4) For a 30" fall, the manual mode delivers 66% of the theoretical free fall energy indicating the method of test does not allow complete free fall.

When looking at the cathead with rope and standard cylinder weight driver data, the engineer looked at three of the factors simultaneously (the three factor interaction was statistically significant). For example shown below uses factors of number of turns of the rope on the cathead (one and two), rope age (old and new rope), and the same fall heights as for Safe-T-Driver -- Safety Hammer combination (27", 30" and 32"). In addition, for comparison of means only, impact velocities for 3 turns of rope on the cathead, old and new rope, for 30" fall are given.

The average (arithmetic) impact velocity and the standard error of approximately 2.5 for the 14 combinations of fall, turns and old and new rope are presented in Table 3.

TABLE 3

Impact Velocity (in./sec.): Fall x Turns x Rope

No. of Turns	Rope Age	Fall		
		27"	30"	32"
1	Old	122	127	133
	New	119	126	130
2	Old	110	117	116
	New	113	122	122
3	Old		103	
	New		92	
			±2.5	

The data in Table 3 is plotted in Figure 2 which shows the "nonparallelism" indicated by the significant three factor interaction (Fall x Turns x Rope).

Three conclusions an engineer may draw from Table 3 and Figure 2 are (Using 2 x 2.5 = 5 as the minimum required difference between means to show significance at the 5% level.):

1) The number of turns of rope around the cathead (used to lift the 140 pound cylinder weight) significantly influences the impact velocity of the weight on the drill stem striker plate and thus the amount of kinetic energy delivered.
2) The influence of rope age on the impact velocity increases with the number of turns of rope used around the cathead. The reasons for these effects are discussed elsewhere (5). With one turn of the rope, the difference is insignificant (for the 30" fall which is used as the standard fall). Whereas, for two or three turns, it becomes an important factor on the impact velocity.
3) The impact velocity increases as the fall height increases (as expected) but not at the rate as anticipated. For example, the velocity of a freely falling weight is 144", 152" and 157" per second for a fall of 27", 30" and 32", respectively.

References

(1) McLean, R. A., Anderson, V. L., Bayer, H. S. and McElrath, G. W., "A Scientific Approach to Experimentation for Consulting Statisticians," Journal of Quality Technology, Vol. 5, No. 1, 1973, pp. 1-6.

(2) Anderson, V. L. and McLean, R. A., Design of Experiments: A Realistic Approach, Marcel Dekker, Inc., New York, 1974.

(3) Kepner, C. H. and Tregoe, The Rational Manager, McGraw-Hill Book Company New York, 1965.

(4) Anderson, V. L., "Restriction Errors for Linear Models (An Aid to Develop Models for Designed Experiments)", Biometrics, Vol. 26, No. 2, 1970.

(5) Kovacs, W. D., Evans, J. C., and Griffith, A. H., "A Comparative Investigation of the Mobile Drilling Company's Manually Operated Safety Driver With the Standard Cathead with Manila Rope for the Performance of the Standard Penetration Test", Report, School of Civil Engineering, Purdue University, W. Lafayette, Indiana, In preparation.