Experimental Designs

Y520
Strategies for Educational Inquiry

Robert S. Michael

Research Methodology

- Is concerned with how the design is implemented and how the research is carried out. The methodology used often determines the quality of the data set generated. Methodology specifies:
  - When and how often to collect data
  - Construction of data collection measures
  - Identification of the sample or test population
  - Choice of strategy for contacting subjects
  - Selection of statistical tools
  - Presentation of findings
Why use experimental designs?

Man prefers to believe what he prefers to be true.

— Francis Bacon

The best method — indeed the only fully compelling method — of establishing causation is to conduct a carefully designed experiment in which the effects of possible lurking variables are controlled. To experiment means to actively change $x$ and to observe the response in $y$.

Why use experimental designs?

The experimental method is the only method of research that can truly test hypotheses concerning cause-and-effect relationships. It represents the most valid approach to the solution of educational problems, both practical and theoretical, and to the advancement of education as a science.


Why use experimental designs?

To propose that poor design can be corrected by subtle [statistical] analysis techniques is contrary to good scientific thinking.

— Pocock, Stuart (19xx). *Controlled clinical trials.* (p.58).
Why use experimental designs?

Issues of design always trump issues of analysis.

— G. E. Dallal, 1999, explaining why it would be wasted effort to focus on the analysis of data from a study whose design was fatally flawed. (http://www.tufts.edu/~gdallal/study.htm)

Why use experimental designs?

100% of all disasters are failures of design, not analysis.

Unique Features of Experiments

- Investigator manipulates a variable directly (the independent variable), while controlling all other potentially influential variables.

- Empirical observations from experiments provide the strongest basis for inferring causal relationships.

Additional Features of Experiments (a)

- Problem statement ⇔ theory ⇔ constructs ⇔ operational definitions ⇔ variables ⇔ hypotheses.

- The research question (hypothesis) is often stated as an alternative to the null hypothesis.

- Random sampling of subjects (insures sample is representative of population)
Additional Features of Experiments (b)

- Random assignment of subjects to treatment and comparison groups (insures equivalency of groups; i.e., unknown variables that may influence outcome are distributed randomly across groups).

- After treatment, performance (dependent variable) of subjects in both groups is compared.

Additional Features of Experiments (c)

- Whenever an experimental design is used, the researcher is interested in determining cause and effect.

- The causal variable is the independent variable and the outcome or effect is the dependent variable.
Independent Variable Manipulation (a)

- **Presence or Absence technique.**
  - An independent variable can be manipulated by presenting a condition or treatment to one group (treatment or experimental) and withholding the condition or treatment from another group (comparison or control).

Independent Variable Manipulation (b)

- **Amount technique.**
  - The independent variable can be manipulated by varying the amount of a condition, such as varying the amount of practice a child is permitted in learning a complex skill (such as learning to play the piano).
Independent Variable Manipulation (c)

- Type of Treatment technique.
  - The independent variable can be manipulated by varying the type of a condition or treatment administered. For example, students in one group outline reading material. Students in another group write a summary sentence for each paragraph. Students in a third group draw concept maps.

Controlling Extraneous Variables (a)

- The goal: All groups (treatment and comparison) are equivalent to each other on all characteristics or variables at the beginning of an experiment.
- The only systematic difference between the groups in an experiment is the presentation of the independent variable. The groups should be the equivalent on all other variables.
- Suppose an investigator wanted to test the effect of a new method of spatial training. If one group is mostly females and the other group is mostly males, then the variable of gender may differentially affect the outcome.
Controlling Extraneous Variables (b)

- Confounding variables (such as gender) are controlled by using one or more procedures that eliminate the differential influence of an extraneous variable.
- **Random assignment of subjects to groups.** This is the best means of controlling extraneous variables in experimental research. Extraneous variables are assumed to be distributed randomly across groups.
- Random assignment controls for both known and unknown confounding variables.

Controlling Extraneous Variables (c)

- Random assignment results in groups that are equivalent on all variables at the start of the experiment.
- Subjects should be randomly selected from a population and then randomly assigned to groups.
- Random selection ensures external validity.
- Random assignment ensures internal validity.
Controlling Extraneous Variables (d)

- Random assignment is more important than random selection because the goal of an experiment is to establish firm evidence of a cause and effect relationship.
- Random assignment controls for the differential influence that confounding extraneous variables may have in a study by insuring that each participant has an equal chance of being assigned to each group.
- In addition, the equal probability of assignment means the characteristics of participants are also equally likely to be assigned to each group.
- Participants and their characteristics should be distributed approximately equally in all groups.

Controlling Extraneous Variables (e)

- Extraneous variables may exist even after random sampling and random assignment (i.e., the groups are not equivalent):
  - Subject mortality (dropouts due to treatment)
  - Hawthorne effect
  - Fidelity of treatment (manipulation check)
  - Data collector bias (double blind studies)
  - Location, history
Controlling Extraneous Variables (f)

- Additional procedures for controlling extraneous variables (used as needed):
  - Use subjects as own control
  - Matching subjects on certain characteristics
  - Blocking
  - Analysis of covariance

Controlling Extraneous Variables (g)

- Matching
  - Match participants in the various groups on the extraneous variable that needs to be controlled.
  - This eliminates any differential influence of that variable.
  - Procedure: Identify the confounding extraneous variables to be controlled, measure all participants on this variable, find another person who has the same amount or type of these variables, and then assign these two individuals, one to the treatment group and one to the comparison group.
Controlling Extraneous Variables (h)

- Hold the extraneous variable constant
  - Controls for confounding extraneous variables by insuring that the participants in the different groups have the same amount or type of a variable.
  - For example, in a study of academic achievement, participants are limited to people who have an IQ greater than 120.

Controlling Extraneous Variables (i)

- Blocking
  - Controls for a confounding extraneous variable by making it an independent variable.
  - For example, in a study of academic achievement, participants in one (comparison and treatment) group consists only of people who have an IQ greater than 120. In another block of comparison and treatment groups, are participants who have IQ's between 80 and 120. And so on.
Controlling Extraneous Variables (i)

- Counterbalancing
  - Used to control for order and carry-over effects. Is relevant only for a design in which participants receive more than one treatment (e.g., repeated measures).
  - Order effects: Sequencing effects that arise from the order in which people take the various treatment conditions.
  - Carry-over effects: Sequencing effects that occur when the effect of one treatment condition carries over to a second treatment condition.

Controlling Extraneous Variables (j)

- Counterbalancing
  - Controls for order and carry-over effects by administering each experimental treatment condition to all groups of participants but in different orders.
Controlling Extraneous Variables (k)

- Analysis of Covariance
  - This is a statistical technique used when participants in various groups differ on some variable that could affect their performance.
  - Example, If students with higher GRE Verbal scores were in one of two groups, these students would be expected to learn faster. Thus you would want to control for verbal skill (assuming that is what GRE Verbal represents).

Controlling Extraneous Variables (l)

- Analysis of Covariance
  - Analysis of covariance statistically adjusts the dependent variable scores for the differences that exist in an extraneous variable.
  - The only relevant extraneous variables are those that also effect participants’ responses to the dependent variable.
Experimental Designs

- Pre-Test / Post-test comparison group
- Post-test only comparison group
- Solomon Four-Group

Randomized, Pre-test / Post-test, Control Group Design

<table>
<thead>
<tr>
<th>Treatment</th>
<th>R</th>
<th>O₁</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>R</td>
<td>O₁</td>
<td>O₂</td>
<td></td>
</tr>
</tbody>
</table>

R = Random assignment
X = Treatment occurs for X₁ only
O₁ = Observation (Pre-test)
O₂ = Observation (Post-test, dependent variable)
(Random sampling assumed)
Pre-test / Post-test control group design

- Subjects (aka, participants) are randomly assigned to the experimental and comparison groups.
- Both groups are pre-tested on the dependent variable, the treatment (aka, independent variable) is administered to the experimental group, and both groups are then post-tested on the dependent variable.

This is a strong design because

- subjects are randomly assigned to groups, insuring equivalence;
- a comparison group is used;
- most of the potentially confounding variables that might threaten internal validity are controlled.
Pre-test / Post-test control group design

Controls:
- Random sampling from population.
- Random assignment of Ss to experimental & comparison groups.
- Timing of independent variable.
- All other variables / conditions that might influence the outcome are controlled.

Control: Random Sampling from Population

- Guarantees all Ss equally likely to selected.
- May assume the sample is representative of the population on all important dimensions.
- No systematic differences between sample and population.
Control: Random Assignment to Groups

- Occurs after random sampling of subjects.
- Guarantees all Ss equally likely to be in comparison or experimental group.
- May assume the two groups are equivalent on all important dimensions.
- No systematic differences between groups.
- May substitute matching on characteristics that might affect dependent variable (e.g., IQ, income, age).
- May not know which characteristics to match on—compromises internal validity.

Control: Independent Variable

- Both groups experience the same conditions, except the experimental group is exposed to independent variable, in addition to the shared conditions.
- Experimenter controls when and how the independent variable is applied to experimental group.
Summary:
Steps in Pre-Test / Post-test Controlled Exp.

- Random sampling of subjects.
- Random assignment of subjects to comparison or experimental groups.
- Administer pre-test to all subjects in both groups.
- Ensure both groups have same conditions, except that experimental group receives the treatment.
- Administer post-test to all subjects in both groups.
- Assess change in dependent variable from pre-test to post-test for each group.

Randomized, Pretest / Post-test, Control Group Design

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>O₁</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R = Random assignment
X = Treatment occurs for X₁ only
O₁ = Observation (Pre-test)
O₂ = Observation (Post-test, dependent variable)
Randomized Pretest / Post-test, Control Group Design

<table>
<thead>
<tr>
<th>Random Assignment of Ss to:</th>
<th>$O_1$ = (Pre-test) Of dependent variable</th>
<th>Exposure to Treatment (X) Independ var</th>
<th>$O_2$ = (Post-test) Of dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group</td>
<td>Exp group’s average score on dep var</td>
<td>X</td>
<td>Exp group’s average score on dep var</td>
</tr>
<tr>
<td>Comparison Group</td>
<td>Comparison group’s average score on dep var</td>
<td></td>
<td>Comparison group’s average score on dep var</td>
</tr>
</tbody>
</table>

Randomized Pretest / Post-test: Within Group Differences

<table>
<thead>
<tr>
<th>Control group Pre-test Score</th>
<th>Control group Post-test Score</th>
<th>= Control group difference on dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group Pre-test Score</td>
<td>Experimental group Post-test Score</td>
<td>= Experimental group difference on dependent variable</td>
</tr>
</tbody>
</table>
Randomized Pretest / Post-test: 
Within Group Differences (b)

The difference in the control group’s score from the pre- to post-test shows the amount of change in the dependent variable that is expected to occur without exposure to the independent or treatment variable.

The difference in the experimental group’s score from pre- to post-test shows the amount of change in the dependent variable that could be expected to occur with exposure to the independent or treatment variable.

Randomized Pre-test / Post-test: 
Between Group Differences

\[
\text{Control group difference} - \text{Experimental group difference} = \text{Difference attributable to independent variable}
\]

The difference between the change in the experimental group and the change in the control group is the amount of change in the dependent variable that can be attributed to the influence of the independent or treatment variable.
Example:
Pre-Test / Post-test Control Group Design

- An educator wants to compare the effectiveness of computer software that teaches reading with that of a standard reading curriculum.
- She tests the reading skill of a group of 60 fifth graders, and then randomly assigns them to two classes of 30 students each.
- One class uses the computer reading program while the other studies a standard curriculum.
- In June she retests the students and compares the average increase in reading skill in each of the two classes.

Example:
Pre-test / Post-test Control Group Design

<table>
<thead>
<tr>
<th>Random Assignment of Ss to:</th>
<th>Score on test of reading skill</th>
<th>Exposure to Treatment (X) independ var</th>
<th>Score on test of reading skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer (Exp Group)</td>
<td>Average 68</td>
<td>X</td>
<td>Average 91</td>
</tr>
<tr>
<td>Comparison Group</td>
<td>Average score: 89</td>
<td></td>
<td>Average score: 104</td>
</tr>
</tbody>
</table>
Example:
Pre-Test / Post-test Control Group Design

- The comparison group’s average reading score was 89 on the pre-test and 104 on the post-test. The difference between pre- and post-tests is 15 points. This is the amount of change in reading skill that could be expected to occur with the regular curriculum only.

- The experimental group’s average reading score was 68 on the pre-test and 91 on the post-test. The difference between pre- and post-tests is 23 points. The change for the experimental group from pre- to post-test indicates the change in reading skill that could be expected to occur with the computer reading curriculum.

Example:
Pre-Test / Post-test Control Group Design

- The difference between the change in the experimental group (23 points) and the change in the comparison group (15 points) is 8 points. This is the amount of change in reading skill that can be attributed solely to the computer reading program.
Example: Rosenthal (1966)
Pre-Test / Post-test Control Group Design

- Recall this experiment

Post-test only control group design

- This design includes all the same steps as the pre-test/post-test design except that the pre-test is omitted.
- Reasons for omitting pre-testing:
  - Subjects already exposed to the treatment
  - Too expensive or too time-consuming
- With large groups, this design controls most of the same threats to internal & external validity. For small groups, a pre-test is necessary.
- Eliminates the threat of pre-testing
- May decrease experimental mortality by shortening the length of the study.
Randomized Post-test, Control Group Design

<table>
<thead>
<tr>
<th>Treatment</th>
<th>R</th>
<th>X₁</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>R</td>
<td></td>
<td>O</td>
</tr>
</tbody>
</table>

R = Random assignment  
X = Treatment occurs for X₁ only  
O = Observation (dependent variable)  
(Random sampling assumed)

Example:  
Randomized Post-test, Control Group  
- Chase 1976
### Randomized Solomon Four Group Design

<table>
<thead>
<tr>
<th>Treatment</th>
<th>R</th>
<th>O₁</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>R</td>
<td>O₁</td>
<td>X₂</td>
<td>O₂</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>R</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>R</td>
<td>X₂</td>
<td>O₂</td>
</tr>
</tbody>
</table>

- R = Random assignment
- X = Treatment occurs for X₁ only
- O₁ = Observation (Pre-test)
- O₂ = Observation (Post-test, dependent variable)

### Randomized Assignment with Matching

<table>
<thead>
<tr>
<th>Treatment</th>
<th>M,R</th>
<th>X₁</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>M,R</td>
<td>O</td>
<td></td>
</tr>
</tbody>
</table>

- M = Matched Subjects
- R = Random assignment of matched pairs
- X = Treatment (for X₁ only)
- O = Observation (dependent variable)
Example:
Randomized Assignment with Matching

- An experimenter wants to test the impact of a novel instructional program in formal logic. He infers from previous research that high ability students and those with programming, mathematical, or musical backgrounds are likely to excel in formal logic regardless of type of instruction.
- Experimenter randomly samples subjects, looks at subjects SAT scores, matches subjects on basis of SAT scores, and randomly assigns matched pairs (one of each pair to each group).
- Other concomitant variables (previous programming, mathematical, and musical experience) could also be matched.
- Difficult to match on more than a few variables.

Randomized Pretest – Post-test
Control Group, Matched Ss

<table>
<thead>
<tr>
<th>Treatment</th>
<th>O₁</th>
<th>M,R</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>O₁</td>
<td>M,R</td>
<td>O₂</td>
<td></td>
</tr>
</tbody>
</table>

- O₁ = Pretest
- M = Matched Subjects
- R = Random assignment of matched pairs
- X = Treatment (for X₁ only)
- O₂ = Observation (dependent variable)
Methods of Matching

- Mechanical
- Statistical

Mechanical Matching

- Rank order subjects on variable(s), take top two, randomly assign members of pair to groups. Repeat for all pairs.
- Problems:
  - Impossible to match on more than one or two variables simultaneously.
  - May need to eliminate some Ss because there is no appropriate match for one of the groups.
Statistical Matching

- Used to control for factors that cannot be randomized by nonetheless can be measured on an interval scale.
- Achieved by measuring one or more concomitant variable (the “covariate”) in addition to the variable (“variate”) of interest (the dependent variable).
- Can be used in experimental, quasi-experimental, and non-experimental designs.

Factorial Designs (a)

- To compare the effects of two or more independent variables in the same experiment, cross them.
- Factorial designs enable the researcher to determine the independent influence of each independent variable, and the interactive influence of the independent variables.
- Each combination of independent variables is called a “cell.”
- Subjects are randomly assigned to as many groups as there are cells.
Factorial Designs (b)

- Groups of subjects are administered the combination of independent variables that corresponds to their assigned cell; then their performance is measured.
- Data collected from factorial designs yields information on the effect of each independent variable separately and the interaction between (and/or among) the independent variables.

Factorial Designs (c)

- Main effect: The effect of each independent variable on the dependent variable.
  - Each independent variable may have a main effect.
  - If a study included the independent variables “type of instruction” and “student verbal skill,” then, potentially, there are two main effects, one for instruction type, and one for verbal skill.
Factorial Designs

Interaction:
- Occurs between two or more independent variables when the effect that one independent variable has on the dependent variable depends on the level of the other independent variable.
- If gender is an independent variable and method of teaching mathematics is another independent variable, then an interaction exists if:
  - the lecture method was more effective for teaching males and
  - individualized instruction was more effective in teaching females mathematics.

Example:
2 x 2 Factorial Design

<table>
<thead>
<tr>
<th>Cells = Conditions = Groups</th>
<th>Reinforcement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
</tr>
<tr>
<td>Ses</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
</tr>
</tbody>
</table>
Example: 2 x 2 Factorial Design  (g)

Results

<table>
<thead>
<tr>
<th>Ses</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>M_{11} = 40</td>
<td>M_{21} = 15</td>
</tr>
<tr>
<td>Middle</td>
<td>M_{12} = 40</td>
<td>M_{22} = 40</td>
</tr>
</tbody>
</table>

Example: 2 x 2 Factorial Design  (h)

Interaction

Mean Vocabulary Score

Immediate

Delayed

Low    Middle

SES
Repeated Measures Design

- In the simplest repeated measures design, each participant receives all treatment conditions but they receive different sequences of treatments (usually randomly determined).
- Each participant is compared with him or herself. Even subtle treatment effects may be statistically significant. Hence, the main advantage of this design is power – the ability to detect small differences.
- The main problem is detecting and dealing with order effects (practice effects, fatigue effects, carryover effects, and sensitization).
- Counterbalancing: used to balance out routine order effects.

Repeated Measures Design

- Suppose a researcher were investigating the effect of type of instruction on learning mathematics.
- Lecture and individualized instruction were the two types.
- All participants would receive both types of instruction but some participants would receive the lecture first and then individualized instruction, while other participants would receive the reverse order.
Counterbalancing: Repeated Measures Design

This design permits us to answer:

- Is the lecture method more effective than the individual instruction? (Main effect of within-subjects factor of type of instruction)

- Do students learn more if they experience the lecture first and then the individualized instruction, or the inverse? (Main effect of the between subjects factor of counterbalancing sequence)

- Is performance better after the second type of instruction than the first? (Instruction by counterbalancing interaction)