6. Experimental, Quasi-Experimental, and Single-Subject Designs

This chapter focuses on four major categories of experimental designs that are used to make causal inferences about the relationship between independent and dependent variables:

- Pre-experimental
- Randomized experimental
- Quasi-experimental
- Single-subject designs

Four criteria for making causal conclusions.

1. ***Causal Relativity.*** In education it is rarely the case that an intervention is compared to a control group that received no intervention. Almost always, choices are made about which of several alternative interventions are most effective. It is not simply a matter of showing that an intervention is effective; the relative causality of one intervention needs to be compared to that of other interventions.
2. ***Causal Manipulation.*** The independent variable that is used to make causal conclusions must be manipulated, or controlled, by the experimenter, unlike assigned independent variables such as age or gender.
3. ***Temporal Ordering.*** There must be a specified time interval during which the intervention is administered. That is, there is an ordering of aspects of the study that occur in sequence—pre intervention, intervention, and post intervention. The intervention is something that is planned ahead to follow from what is known before the intervention.
4. ***Elimination of Alternative Explanations.*** Good experiments are designed so that there are no plausible rival explanations of what caused post intervention differences. Outcomes need to be attributed to the intervention and not to other variables or influences.

Characteristics of Experimental Research

Six distinguishing characteristics of traditionally defined experimental research ;

1. ***theory-driven research hypotheses***
   Having theory-driven research hypotheses explains why there are cause-and-effect findings and provides a clear indication of how the findings should be generalized.
2. ***statistical equivalence of subjects in intervention and control and/or comparison groups (achieved through random assignment)***
   Achieving statistical equivalence of individuals in groups that are compared is essential for ruling out the many possible variables that could invalidate causal conclusions.
3. ***researcher-controlled interventions independently and uniformly applied to all subjects***
   Researcher-controlled interventions, or direct manipulation of the intervention, is perhaps the most distinct feature of experimental research.
4. ***measurement of each dependent variable***
   Measurement of dependent variables means that experimental research is concerned with outcomes that can be assigned a numerical value.
5. ***use of inferential statistics***
   Inferential statistics are used to make probability statements about the results.
6. ***rigorous control of conditions and extraneous variables***
   We control extraneous variables either by making sure that they have no effect on the dependent variable or by keeping the effect the same for all groups.

Ideal and actual characteristics of strong and weak evidence in experimental research; (p281)

Ideal—Leading to Strong Evidence	Actual—Leading to Weak Evidence
Restricted, controlled environment	Complex, uncontrolled environment
Results clearly attributed to intervention	Factors other than the intervention may affect results
Random assignment of subjects	Subjects rarely randomized
Sufficiently large sample	Due to nesting within classes and schools, small unit of analysis
Completely separate experimental and control groups	Interaction between experimental and control groups
Tightly controlled administration of the intervention	Messy, difficult-to-control implementation
Conducted with sufficient time to allow intervention effects	Tends to be conducted over shorter periods of time
Experimental and control group environments are the same	Experimental and control group environments are different

The first steps in planning experimental research are to define a research problem, search the literature, and state clear research hypotheses.

Next, the researcher selects subjects from a defined population and, depending on the specific design used, usually assigns subjects to different groups. A simple experimental study involves two groups, one called the intervention, experimental, or treatment group and the other called the control or comparison group.

- **Intervention, experimental, or treatment group**
  : subjects who receive the intervention

- **Control or comparison group**
  : subjects who do not receive the targeted intervention
  or receive a different intervention

It is most common to conceive of the two groups in experimental research as the treatment and comparison groups or as a design with two interventions.

One of the difficulties in planning experimental research is knowing whether the interventions will be strong enough.

Another important consideration in designing an experiment is to be sure that the treatment or intervention has occurred as planned, which is called **fidelity of intervention**.

To determine if the intervention was implemented as designed, five criteria can be used (O'Donnell, 2008):

1. ***Adherence***—whether each component of the intervention is delivered as designed
2. ***Duration***—whether the intervention was implemented with a sufficient length and number of sessions
3. ***Quality of delivery***—whether the techniques, processes, and procedures as prescribed are delivered
4. ***Participant responsiveness***—whether participants are engaged in and involved with program activities and content
5. ***Program differentiation***—whether features that distinguish the intervention from other programs are present

Statistical Conclusion Validity

- Meaning : the correctness of statistical analyses (“Statistical Power”)
- Summary of threats to statistical conclusion validity (p284, TABLE 11.3)

Threat	Description
Low statistical power	The design does not have enough subjects or a powerful enough intervention to detect a difference.
Violated assumptions of statistical tests	Assumptions such as having a population with a normal distribution and equal variances are not met, leading to incorrect support or incorrect nonsupport of the research hypotheses.
“Fishing” and the error rate problem	A statistically significant difference has been found with one of many statistical tests on the same data.
Unreliability of measures	The presence of measurement error makes it difficult to obtain a significant difference.
Restriction of range	Small variances or ranges make it difficult to obtain significant relationships.
Unreliability of intervention implementation	Differences in the administration of an intervention to different individuals or groups result in underestimating the effect of the intervention.
Extraneous variance in the experimental setting	Differences in the settings in which the interventions took place inflate the error rate, making it more difficult to find a relationship or difference.

Internal Validity

- Meaning: Causal truthfulness
- The internal validity of a study is a judgment that is made concerning the confidence with which plausible rival hypotheses can be ruled out as explanations for the results.
- Summary of Threats to Internal Validity (p285, TABLE 11.4)

Threat	Description
History	Unplanned or extraneous events that occur during the research may affect the results.
Selection	Differences between the subjects in the groups may result in outcomes that are different because of group composition.
Statistical regression	Scores of groups of subjects take on values closer to the mean due to respondents' being identified on the basis of extremely high or low scores.
Pretesting	The act of taking a test or responding to a questionnaire prior to the treatment affects the subjects.
Instrumentation	Differences in results are due to unreliability, changes in the measuring instrument, or observers.
Attrition	The systematic loss of subjects affects the outcome.
Maturation	An effect is due to maturational or other natural changes in the subjects (e.g., being older, wiser, stronger, tired).
Diffusion of intervention	Subjects in one group learn about interventions or conditions for different groups.
Experimenter effects	Deliberate or unintended effects of the researcher influence subjects' responses.
Intervention replications	Number of replications of the intervention is different from the number of subjects.
Subject effects	Changes in behavior result in response to being a subject or to being in an experiment.

- Meaning: Labeling of inferences
- In experiments, the term construct validity describes how well measured variables and interventions represent the theoretical constructs that have been hypothesized.
- Summary of threats to construct validity(p286, TABLE 11.5)

Threat	Description
Inadequate explication of the constructs	Invalid inferences are made about the constructs because they have not been sufficiently described and supported by theory.
Mono-operation bias	Only a single type of intervention or dependent variable is used when using multiple types would lead to more assurance that the more abstract theory is supported.
Mono-method bias	Implementing the intervention or measuring the dependent variable in only one way restricts inferences to just those methods as well as to the hypothesized theoretical relationships.

External Validity

- Meaning: Generalizability of results
- External validity is the extent to which the results of an experiment can be generalized to people and environmental conditions outside the context of the experiment.
- Summary of threats to external validity(p286, TABLE 11.6)

Threat	Description
Population	
Selection of subjects	Generalization is limited to the subjects in the sample if the subjects are not selected randomly from an identified population.
Characteristics of subjects	Generalization is limited to the characteristics of the sample or population (e.g., socioeconomic status, age, location, ability, race).
Subject/treatment interaction	Generalization may be limited because of the interaction between the subjects and the intervention (i.e., the effect of the intervention is unique to the subjects).
Ecological	
Description of variables	Generalization is limited to the operational definitions of the independent and dependent variables.
Multiple-treatment interference	In experiments in which subjects receive more than one intervention, generalizability is limited to similar multiple-intervention situations because of the effect of the first intervention on subsequent treatments.
Setting/treatment interaction	Generalization is limited to the setting in which the study is conducted (e.g., room, time of day, others present, other surroundings).
Time of measurement/ treatment interaction	Results may be limited to the time frame in which they were obtained. Interventions causing immediate effects may not have lasting effects.
Pretest-posttest sensitization	The pretest or posttest may interact with the intervention so that similar results are obtained only when the testing conditions are present.
Novelty or disruption effect	Subjects may respond differently because of a change in routine, and generalization may be limited to situations that involve similar novelty or disruption (e.g., an initially effective intervention may become ineffective in time as the novelty wears off).

SINGLE-FACTOR PRE-EXPERIMENTAL DESIGNS

The three designs summarized in this section are termed pre-experimental designs because they are without two or more of the six characteristics of experimental research listed earlier. As a consequence, few threats to internal validity are controlled.

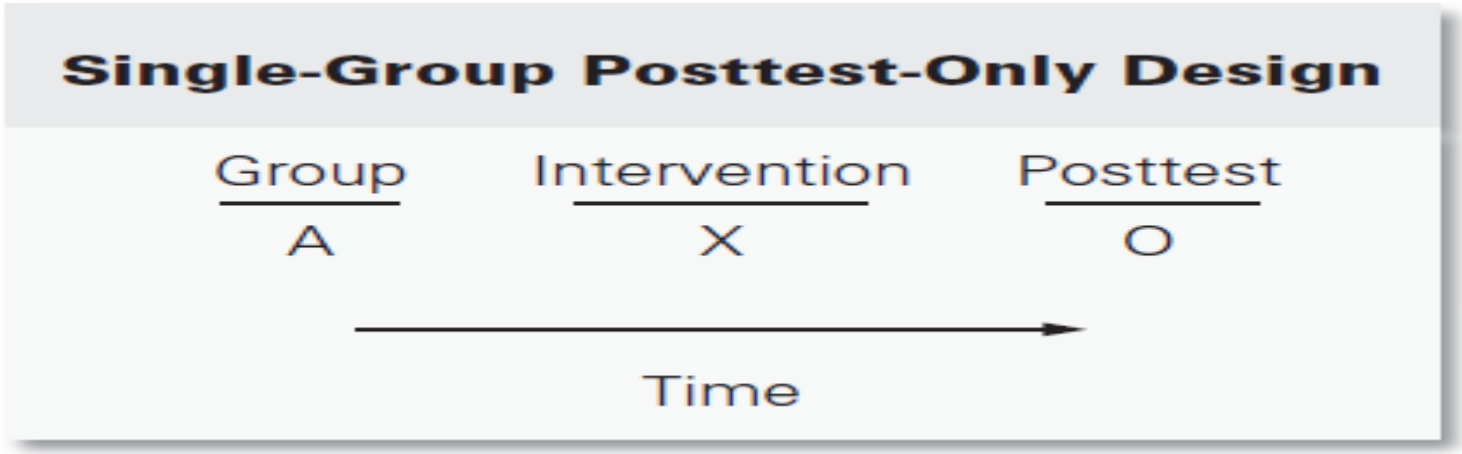
1. Single-Group Posttest-Only Design
2. Single-Group Pretest-Posttest Design
3. Nonequivalent Groups Posttest-Only Design

We will use a notational system to provide information for understanding the designs.

R Random assignment
O Observation, a measure that records pretest or posttest scores
X Intervention conditions
(subscripts 1 through n indicate different interventions)

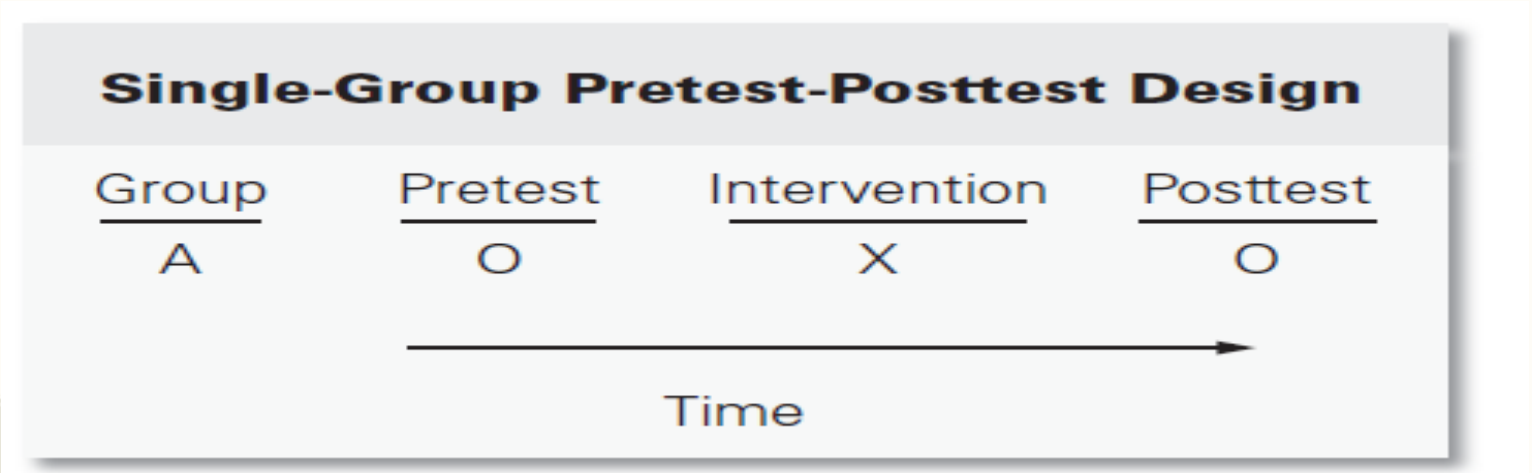
A, B, C, D, E, F Groups of subjects or, for single-subject designs, baseline or treatment conditions

1. Single-group posttest-only design: no comparison group or pretest



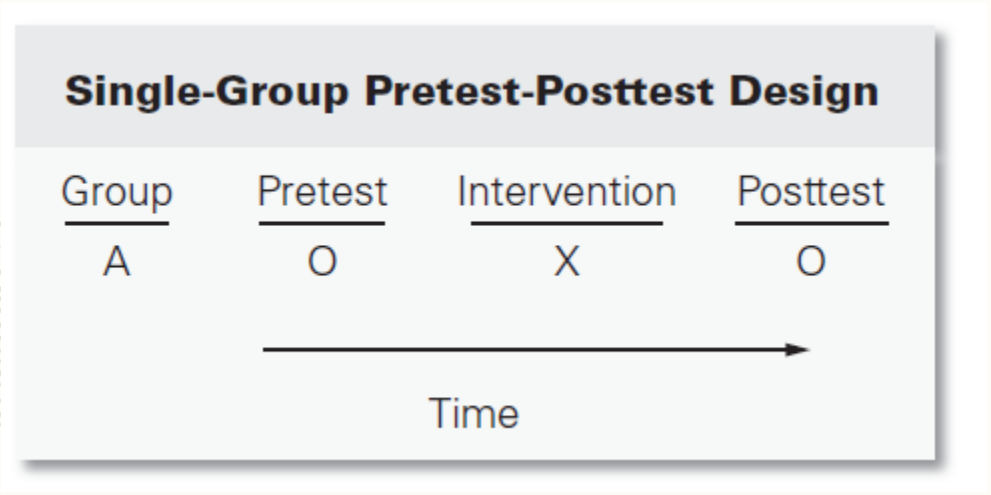
- Although not all threats to internal validity are applicable to this design because there is no pretest and no comparison with other treatments, valid causal conclusions are rare.
- Without a comparison or control group, it is also difficult to know whether other factors occurring at the same time as the intervention were causally related to the dependent variable.
- The only situation in which this design is reasonable is when the researcher can be fairly certain of the level of knowledge, attitude, or skill of the subjects before the intervention, and can be fairly sure that history is not a threat.

2. Single-group pretest-posttest design: no comparison group with a pretest



- Threats to Internal Validity : history, statistical regression, pretesting, attrition, maturation, intervention replications
- Experimenter effects, subject effects, and statistical conclusion threats are possible in any experiment, and these would need to be examined.
- This design should be used only under certain conditions that minimize the plausibility of the threats and when it is impossible to use other designs that will control some of these threats.

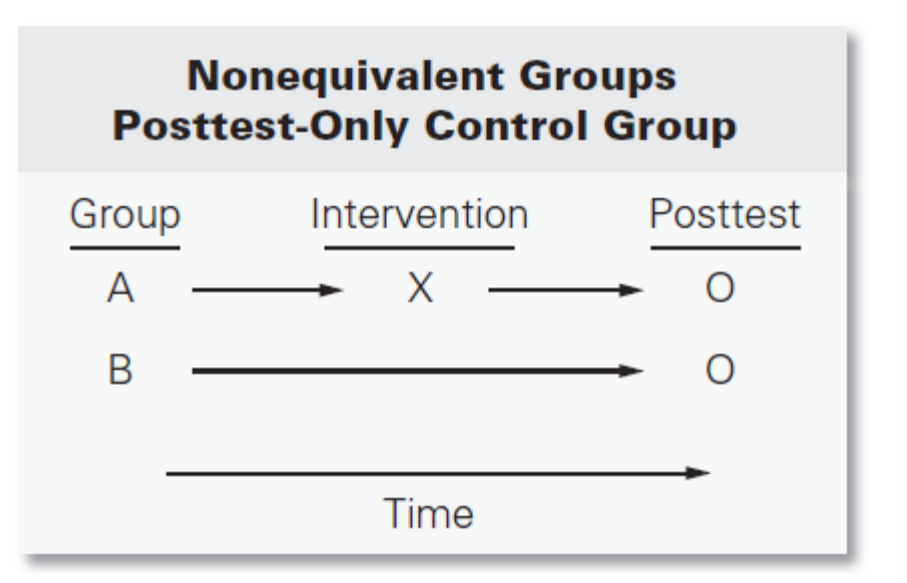
2. Single-group pretest-posttest design: no comparison group with a pretest



- Several modifications can be made to the single-group pretest-posttest design that will improve internal validity, including the following:
  - Adding a second pretest
  - Adding a second pretest and posttest of a construct similar to the one being tested
  - Following the posttest with a second pretest/posttest with the intervention either removed or repeated and determining if the pattern of results is consistent with predictions

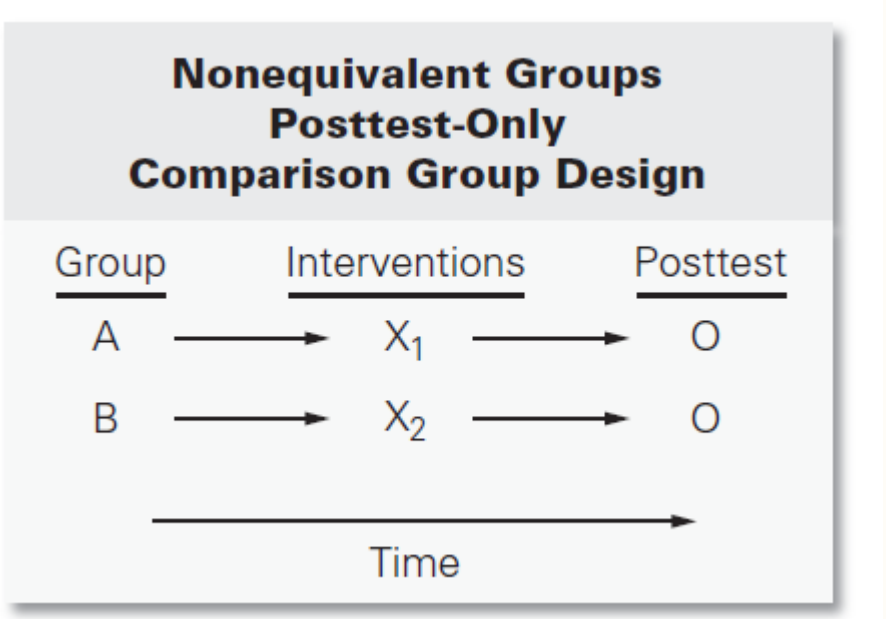
3. Nonequivalent groups posttest-only design: no pretest with a control or comparison group

- Difference from single-group posttest-only design
  - In a nonequivalent groups posttest-only design, a group that receives no intervention or a different intervention is added to the single-group posttest-only design.



- The term *nonequivalent groups* is used for the design because selection is the most serious threat to the internal validity of the results.
- There is no random assignment of subjects to each group.

3. Nonequivalent groups posttest-only design: no pretest with a control or comparison group



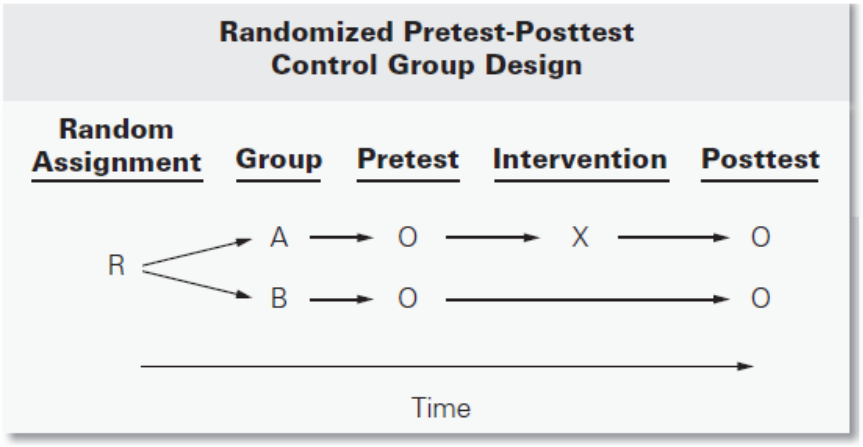
- This design is used when a researcher wants to compare two or more interventions but cannot give a pretest or randomize the assignment of subjects to each group.
- Threats to Internal Validity : internal or within-group history, regression, instrumentation, Attrition, Maturation, experimenter effects, subject effects, intervention replications, statistical conclusion

These include procedures for ruling out group differences through randomization of subjects to groups.

1. ***Randomized Pretest-Posttest Control Group Design***
2. ***Randomized Pretest-Posttest Comparison Group Design***
3. ***Randomized Posttest-Only Control and Comparison Group Designs***

1. Randomized pretest-posttest control group design: random assignment with a control group, pretest, and posttest

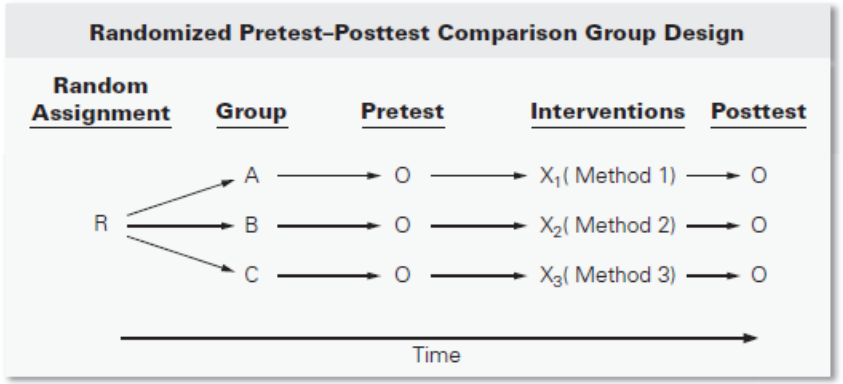
- The purpose of random assignment is to enable the researcher to reasonably rule out any differences between the groups that could account for the results.

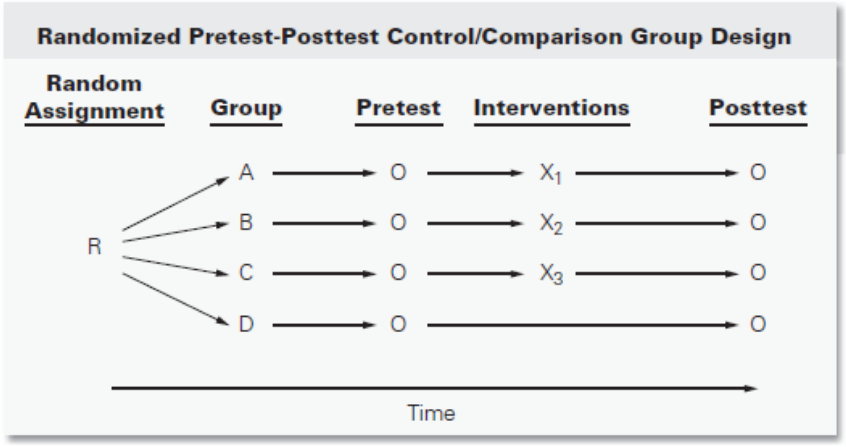


- The first step is random assignment of the subjects to the experimental group and the control group.
- The second step is to pretest each group on the dependent variable.
- The third step is to administer the intervention to the experimental group but not to the control group, keeping all other conditions the same for both groups so that the only difference is the manipulation of the independent variable.

2. Randomized groups pretest-posttest comparison group design:

random assignment with a comparison group, pretest, and posttest



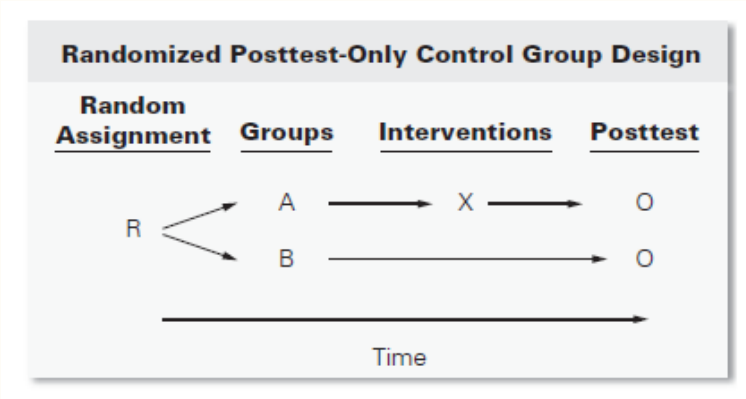


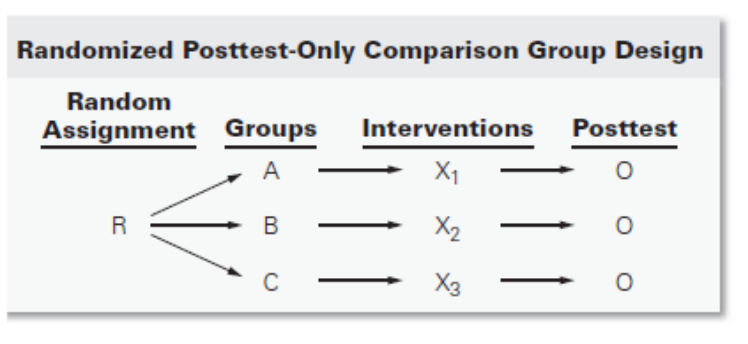
- The pretest-posttest control group design controls four sources of threats to internal validity.- Selection, Statistical Regression, Pretesting, Maturation

3. Randomized posttest-only control and comparison group designs:

random assignment with no pretest

- The randomized posttest-only group design is used when it is unfeasible or inconvenient to give a pretest and in situations in which the pretest might have an effect on the intervention.





3. Randomized posttest-only control and comparison group designs:
random assignment with no pretest

- Four disadvantages to using a randomized posttest-only

1. It is possible that randomization has not controlled for initial group differences or that the lack of a pretest makes it difficult either to check whether differences exist or to control statistically those differences that may be found.
2. The researcher is unable to form subgroups on the basis of the pretest for investigating effects of the intervention on different subgroups.
3. The researcher is unable to determine whether differential attrition has occurred.
4. The statistical analysis is less precise and less likely to show a difference between the groups.

3. Randomized posttest-only control and comparison group designs:
random assignment with no pretest

- The pretest-posttest design may be preferable in these situations:

1. There are subtle, small differences between intervention conditions.
2. Differential mortality is possible.
3. Subgroup analysis is desirable.
4. Anonymity is unnecessary.
5. Pretesting is a normal part of the subjects' routine.

True experimental designs provide the strongest, most convincing arguments of the causal effect of the independent variable. However, true experimental designs cannot be employed because random assignment of subjects to experimental and control groups is impossible and a control or comparison group is unavailable, inconvenient, or too expensive. Fortunately, there are several good designs that can be used under either of these circumstances. They are termed quasi-experimental designs.

1. ***Nonequivalent Groups Pretest-Posttest Control or Comparison Group Designs***
2. ***Time-Series Designs***
   Single-Group Interrupted Time-Series Design
   Control Group Interrupted Time-Series Design

1. ***Nonequivalent groups pretest-posttest control or comparison group designs:*** no random assignment with a pretest and posttest

- Nonequivalent groups pretest-posttest control or comparison group designs are very prevalent and useful in education, because it is often impossible to randomly assign subjects.

Nonequivalent Groups Pretest–Posttest Control Group Design

Group	Pretest	Intervention	Posttest
A	O	X	O
B	O		O

Time

Class	Pretest	Method	Posttest
A	O	X_1	O
B	O	X_2	O
C	O	X_3	O

Time

- The most serious threat to the internal validity : selection

1. ***Nonequivalent groups pretest-posttest control or comparison group designs:***
   no random assignment with a pretest and posttest

- The most serious threat to the internal validity : selection

Consequently, if the researcher knows in advance that randomization is impossible, the groups should be selected to be as similar as possible.
The pretest scores and other measures on the groups are then used to adjust the groups statistically on the factor that is measured.

Another approach to controlling selection when intact groups such as classrooms must be used is to use a large number of groups and then randomly assign entire groups to either control or intervention conditions.

This procedure then changes the study to a true experimental design.

1. ***Nonequivalent groups pretest-posttest control or comparison group designs:***
   no random assignment with a pretest and posttest

- Quasi-experimental designs that use either control or comparison groups can be strengthened by taking the following measures:

  ✓ Adding a second pretest
  ✓ Replicating the treatment with another group at another time with the same pretest
  ✓ Using a comparison group that reverses the effect of the targeted intervention

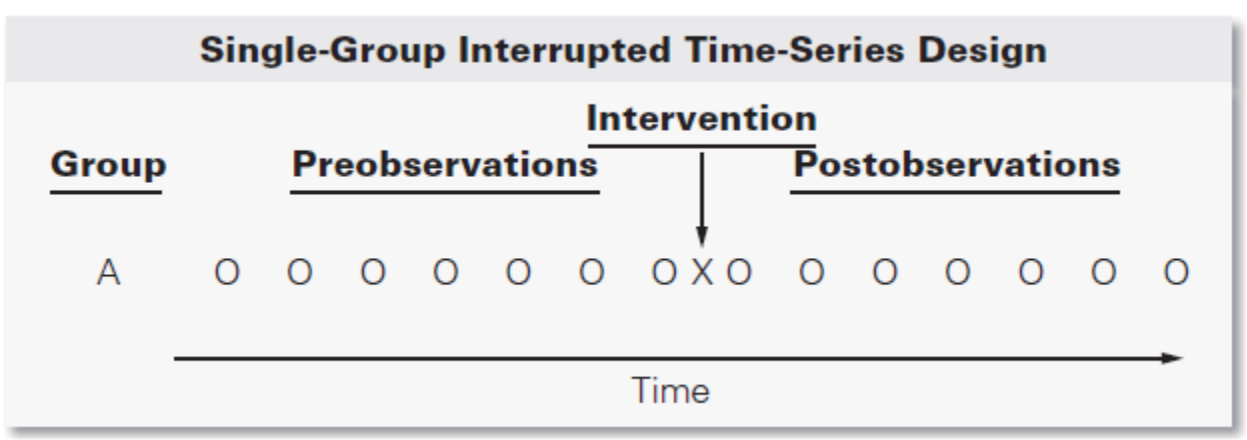
2. Time-series design: intervention with many observations before and after

- If the group is repeatedly measured before and after the intervention, rather than once before and once after, a time-series design or abbreviated time-series design is created. Time-series designs are especially useful when there are continuous, naturally occurring observations of the dependent variable over time and there is a sudden or distinct intervention during the observations.

  ① ***Single-Group Interrupted Time-Series Design***
  ② ***Control Group Interrupted Time-Series Design***

2. Time-series design: intervention with many observations before and after

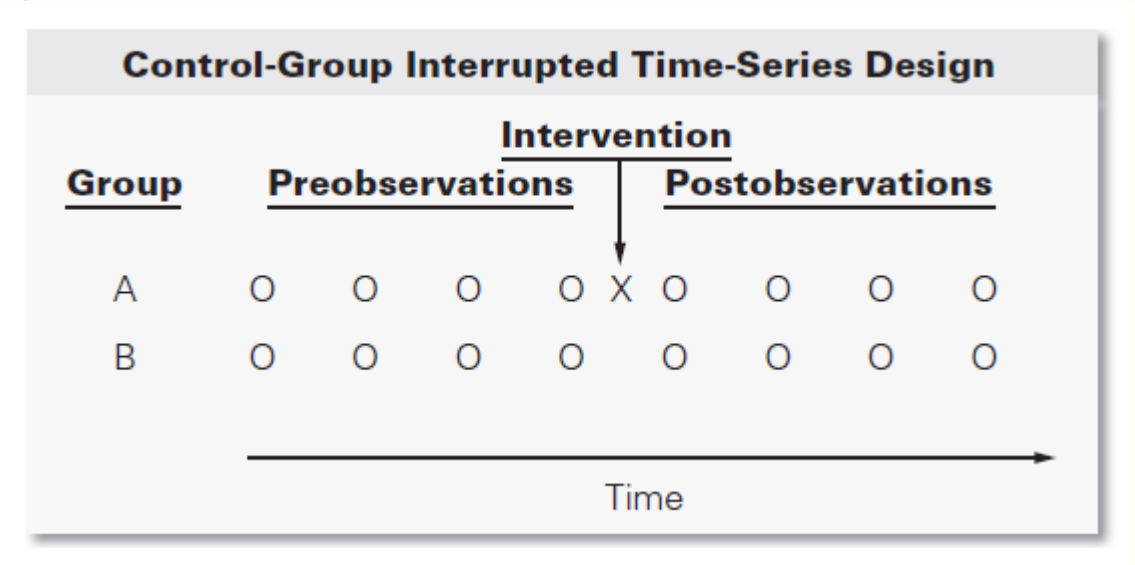
① ***Single-Group Interrupted Time-Series Design :*** one group, one intervention, with many observations before and after



- Several conditions should be met in employing this design.
  1. The observations should be made at equal time intervals and conducted with the same procedures in order to reduce the threat of instrumentation.
  2. The intervention introduced should be a distinctive, abrupt intervention that is clearly new to the existing environment.
  3. There should be some evidence that the subjects involved in each observation are the same.
  4. There should not be any kind of change affecting the subjects occurring at about the same time as the intervention.

2. Time-series design: intervention with many observations before and after

② **Control Group Interrupted Time-Series Design*:*** intervention and control groups



- The important point is that there are weaknesses in all research designs, and it is necessary for the investigator and the reader of research to search out and analyze plausible rival hypotheses that may explain the results.

THREATS TO INTERNAL VALIDITY OF EACH EXPERIMENTAL DESINES

TABLE 11.7 Threats to Internal Validity of Pre-Experimental Designs

Design	History	Selection	Statistical Regression	Pretesting	Instrumentation	Attrition	Maturation	Diffusion of Intervention	Experimenter Effects	Intervention Replications	Subject Effects
Single-group posttest only	–	–	?	NA	?	?	–	NA	?	?	?
Single-group pretest-posttest	–	?	?	–	?	?	–	NA	?	?	?
Nonequivalent groups posttest only	?	–	?	NA	?	?	?	?	?	?	?

TABLE 11.8 Threats to Internal Validity of Randomized Experimental Designs

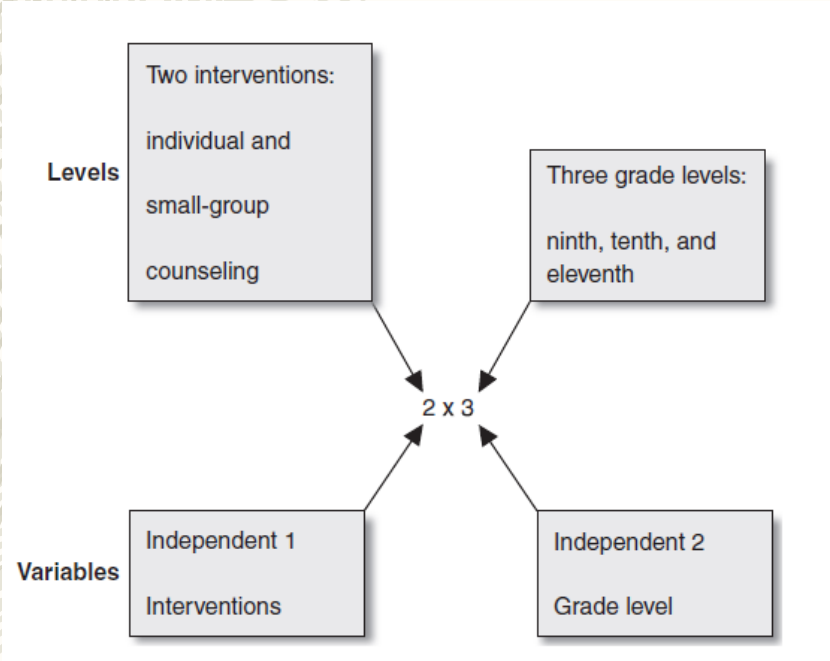
Design	History	Selection	Statistical Regression	Pretesting	Instrumentation	Attrition	Maturation	Diffusion of Intervention	Experimenter Effects	Intervention Replications	Subject Effects
Pretest-posttest control and/or comparison group	?	+	+	+	?	?	+	?	?	?	?
Posttest-only control and/or comparison group	+	+	NA	NA	?	?	+	?	?	?	?

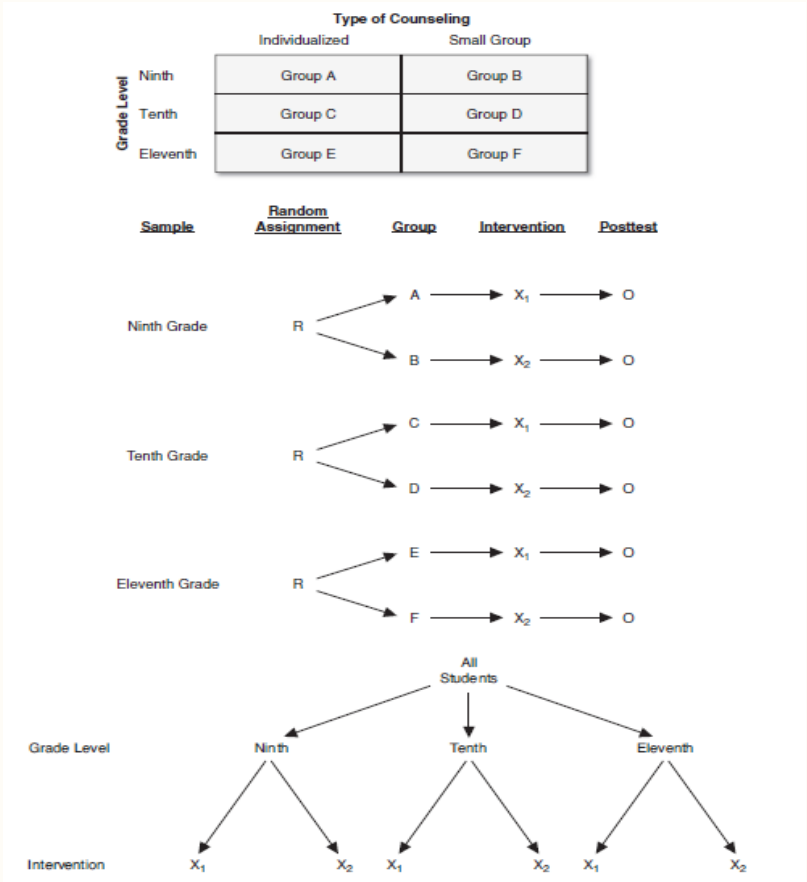
TABLE 11.9 Threats to Internal Validity of Quasi-Experimental Designs

Design	History	Selection	Statistical Regression	Pretesting	Instrumentation	Attrition	Maturation	Diffusion of Intervention	Experimenter Effects	Intervention Replications	Subject Effects
Nonequivalent groups pretest-posttest design	?	–	?	–	?	?	–	?	?	?	?
Single-group interrupted time-series	–	?	+	?	?	?	+	NA	+	?	?
Control group interrupted time-series	+	?	–	+	?	?	+	?	?	?	?

Factorial designs: two or more independent variables analyzed together

- Factorial designs are used for two primary purposes:
  1. to see if the effects of an intervention are consistent across characteristics of the subjects (e.g., age, aptitude, or gender)
  2. to examine the unique effect of the independent variables together (this is called an interaction)





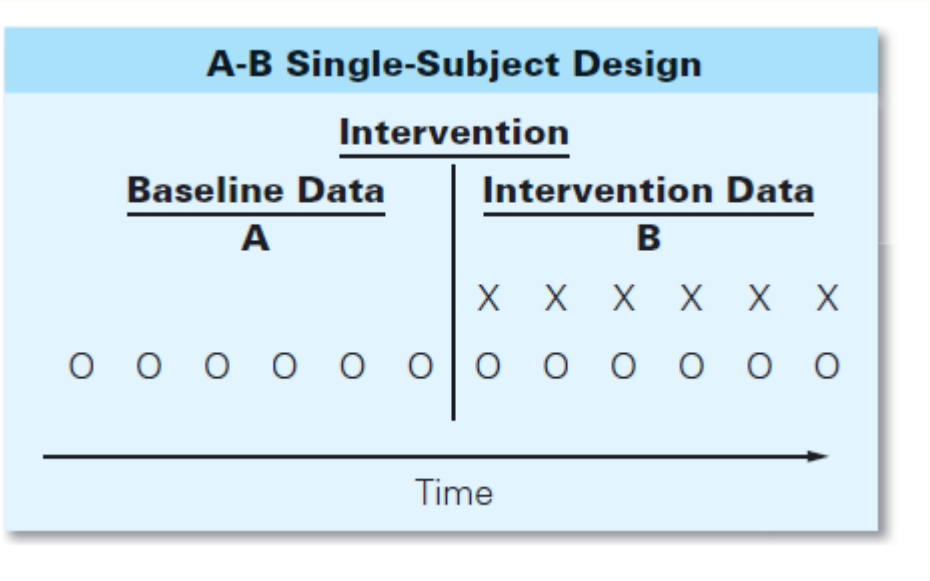
Single-subject designs: one or a few subjects with an intervention

- The most important characteristics of single-subject designs can be summarized as follows:
  Reliable measurement
  Repeated measurement
  Description of conditions
  Baseline and intervention condition; duration and stability
  Single-variable rule

1. A-B Design
2. A-B-A Design
3. Multiple-Baseline Designs

1. A-B design: baseline and intervention comparison

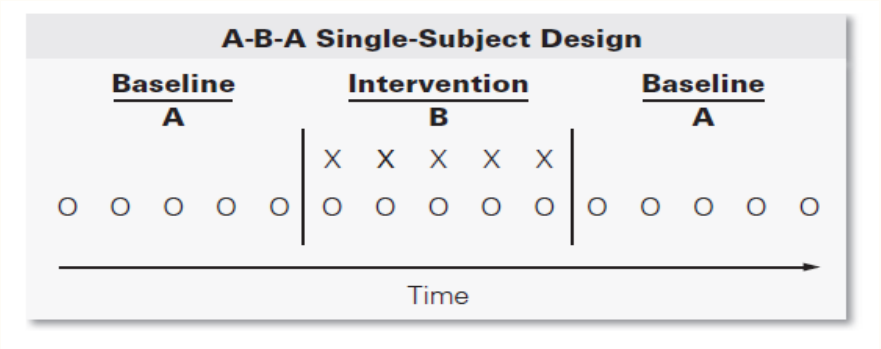
- The A-B design is the most simple and least interpretable single-subject design. The procedure in using it is to observe the target behavior until it occurs at a consistent, stable rate.

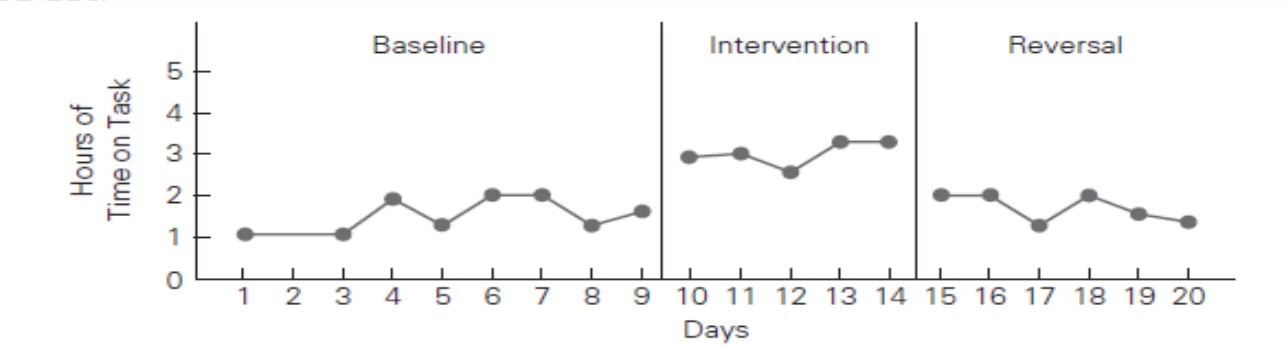


- It is relatively weak in internal validity.

2. A-B-A design: baseline, intervention, and baseline comparison

- Reversal, removal, or withdrawal design: adding an intervention and then taking it away.
- This design allows strong causal inference if the pattern of behavior changes during the intervention phase and then returns to about the same pattern as observed in the first baseline after the intervention is removed.





3. Multiple-baseline designs: more than one subject, variable, or context

- When it is impossible or undesirable to remove an intervention condition or when the effects of an intervention extend into a second baseline phase, strong causal inference can be made by using multiple-baseline designs rather than a simple A-B design. Multiple-baseline designs employ the A-B logic, but rather than using one subject and one kind of target behavior, the researcher collects data on two or more actions, subjects, or situations or some combination of actions, situations, and subjects.

  - *Multiple Baselines across Behavior*
  - *Multiple Baselines across Situations*
  - *Multiple Baselines across Individuals*

Randomized Experimental Designs

1. Was the research design described in sufficient detail to allow for replication of the study?
2. Was it clear how statistical equivalence of the groups was achieved? Was there a full description of the specific manner in which subjects were assigned randomly to groups?
3. Was a true experimental design appropriate for the research problem?
4. Was there manipulation of the independent variable?
5. Was there maximum control over extraneous variables and errors of measurement?
6. Was the intervention condition sufficiently different from the comparison condition for a differential effect on the dependent variable to be expected?
7. Were potential threats to internal validity reasonably ruled out or noted and discussed?
8. Was the time frame of the study described?
9. Did the design avoid being too artificial or restricted for adequate external validity?
10. Was an appropriate balance achieved between control of variables and natural conditions?
11. Were appropriate tests of inferential statistics used?

Quasi-Experimental Designs

1. Was the research design described in sufficient detail to allow for replication of the study?
2. Was a true experiment possible?
3. Was it clear how extraneous variables were controlled or ruled out as plausible rival hypotheses?
4. Were all potential threats to internal validity addressed?
5. Were the explanations ruling out plausible rival hypotheses reasonable?
6. Would a different quasi-design have been better?
7. Did the design approach a true experiment as closely as possible?
8. Was there an appropriate balance between control for internal validity and for external validity?
9. Was every effort made to use groups that were as equivalent as possible?
10. If a time-series design was used,
    (a) Was there an adequate number of observations to suggest a pattern of results?
    (b) Was the intervention introduced distinctly at one point in time?
    (c) Was the measurement of the dependent variable consistent?
    (d) Was it clear, if comparison groups were used, how equivalent the groups were?

Single-Subject Designs

1. Was the sample size one or just a few?
2. Was a single-subject design most appropriate, or would a group design have been better?
3. Were the observation conditions standardized?
4. Was the behavior that was observed defined operationally?
5. Was the measurement highly reliable?
6. Were sufficient repeated measures made?
7. Were the conditions in which the study was conducted fully described?
8. Was the baseline condition stable before the treatment was introduced?
9. Was there a difference between the length of time or number of observations between the baseline and intervention conditions?
10. Was only one variable changed during the intervention condition?
11. Were threats to internal and external validity addressed?

Thank you