Topic 4 ▶ Experimental Research Designs

LEARNING OUTCOMES

By the end of this topic, you should be able to:

1. Define what is research designs;
2. Distinguish the ways in which good research designs differ from weak research designs;
3. Explain the differences between a true experimental design and a quasi-experimental design;
4. Elaborate on the concept of main effects and interaction; and
5. Discuss the role of hypothesis testing in an experiment.

▶ INTRODUCTION

What is meant by research design? According to Christensen (1988) "research design refers to the outline, plan, or strategy specifying the procedure to be used in seeking an answer to the research question. It specifies such things as how to collect and analyse the data" (p.219). The design of an experiment will show how extraneous variables are controlled or included in the study (refer to the control techniques discussed in Chapter 3). The design will determine the types of analysis that can be done to answer your research questions and the conclusions that can be drawn. To what extent your design is good or bad will depend on whether you are able to get the answers to your research questions. If your design is faulty, the results of the experiment will also be faulty. How do you about getting a good research design that will provide answers to the questions asked? It is not easy and there is not fixed way of telling others how to do it. The best that can be done is to examine different research designs and to point out their strengths and weaknesses, and leave it to you to make the decision.

You should have an in depth understanding of your research problem; such as the treatment you want to administer, the extraneous variables or factors you want to control and the strengths and weaknesses of the different alternative designs. You should be clear about your research question/s and what is it you intend to establish. You should avoid selecting a design and then trying to fit the research question to the design. It should be the other way round! Most important is to see if the design will enable you to answer the research question. You should be clear what factors you wish to control so that you can arrive at a convincing conclusion. Choose a design that will give you maximum control over variables or factors that explain the results obtained.

4.1 SYMBOLS USED IN EXPERIMENTAL RESEARCH DESIGNS

Research Design be thought of as the structure of research, i.e. it is the 'glue' that holds all of the elements in a research project together. In experimental research, a few selected symbols are used to show the design of a study.

O = Observation or Measurement (e.g. mathematics score, score on an attitude scale, weight of subjects, etc.).

O_1, O_2, O_3 ………… On = more than one observation or measurement.

R = Random assignment: subjects are randomly assigned to the various groups.

X = Treatment which may be a teaching method, counselling techniques, reading strategy, frequency of questioning and so forth.

4.2 WEAK DESIGNS

We will discuss three types of weak designs which are one-shot design, one-group pretest-posttest design and non-equivalent posttest-only.

4.2.1 One-Shot Design

For example, you want to determine whether praising primary school children makes them do better in arithmetic (see Figure 4.1). You measure arithmetic achievement with a test. To test this idea, you choose a class of year 4 pupils and increase praising of children and you find that their mathematics performance significantly improved.

X	O
(praise)	(scores on a mathematics test)

Figure 4.1: One-shot design

You conclude that praising children increases their mathematics score. This design is weak for the following reasons:

1. **Selection Bias:** It is possible that the pupils you selected as subjects were already good in mathematics.
2. **History:** The school had organised a motivation course on mathematics for year 4 pupils. So, it is possible it might influence their performance.

4.2.2 One-Group Pretest-Posttest Design

To ensure that there was no pre-existing characteristic among the school children, a pretest may be administered (see Figure 4.2). If the children performed better in mathematics after praising compared to the pretest, then you can attribute it to the practice of praising.

This design is weak for the following reasons:

1. Maturation: If time between the pretest and posttest is long, it is possible that the subjects may have matured because of developmental changes.
2. Testing: Sometimes the period between the pretest and the posttest is too short and there is the possibility that subjects can remember the questions and answers.

O_1	X	O_2
(Mathematics pretest)	(praise)	(mathematics posttest)

Figure 4.2: One-Group Pretest-Posttest Design

ACTIVITY 4.1

Twenty pupils who were weak in arithmetic were taught arithmetic using the Zandox method. Three weeks later when they were tested, their arithmetic scores improved. Thus the Zandox method improves arithmetic performance.

1. Which type of research design is this study based on?
2. What are some problems with this design?

4.2.3 Non-Equivalent Posttest-Only

The main weakness of the previous two designs is the lack of a comparison group and the consequent difficulty of saying conclusively that the treatment ('praising') contributed to increased mathematics score. In the Non-Equivalent Posttest-Only Design an attempt is made to include a comparison group (i.e. control group) that did not receive 'praise' (see Figure 4.3). The dashed lines separating the experimental group and the control group indicates that the children were not randomly assigned to the two groups. Hence, the two groups are non-equivalent. Matching can be used but there is no assurance that the two groups can be equated (see Chapter 3). The only way one can have assurance that the two groups are equated is to assign the children randomly.

EXPERIMENTAL GROUP	X	O
- - - - - - - - - -	- - - - - - - -	- - - - - - - - - -
		O
CONTROL GROUP	(Praise)	(Mathematics posttest)

Figure 4.3: Non-equivalent posttest only design

This design is weak for the following reason:

1. **Selection Bias:** Since there was no random assignment, it cannot be established that the two groups are equivalent. So, any differences in the posttest may not be attributable to giving praise but other factors such as ability, IQ, interest and forth.

The three designs describes are 'weak' research designs because they do not allow for controlling of extraneous that might creep into the experiment. These extraneous factors may affect the results of the dependent measure. For example, if attitude towards mathematics and outside tuition in mathematics are not controlled it may not possible to conclude that 'praise' (treatment) affects mathematics performance (dependent variable). Also, weak research designs do not attempt to randomly assignment subjects to the groups being compared which introduce extraneous factor affecting the dependent measure. Random assignment controls for both known and unknown extraneous variables that might affect the results of the experiment.

SELF-CHECK 4.1

1. Identify the major differences between the one-shot design, one-group pretest-posttest design and non-equivalent posttest only design.
2. Why these designs are considered weak?

ACTIVITY 4.2

A teacher assigns one class of pupils to be the experimental group and another class the control group. Both groups are given a science posttest. he pupils in the experimental group are taught by their peers while pupils in the control group are taught by their teacher.

1. Which research design is the teacher using?
2. How will you challenge the findings of the experiment?

4.3 TRUE DESIGNS

In this section, we will discuss some "true" experimental research designs. What is a "true" experimental design? According to Christensen (1988), "to be true experimental design, a research design must enable the researcher to maintain control over the situation in terms of assignment of subjects to groups, in terms of who gets the treatment condition, and in terms of the amount of treatment condition that subjects receive" (p. 231).

In this chapter, we will discuss two major types of true designs: 1) after-only design 2) before-after design (see Figure 4.4). What is the difference between the two designs? The after-only design relies only a posttest while the before-after design (as the name suggests) relies on both a pretest and a posttest.

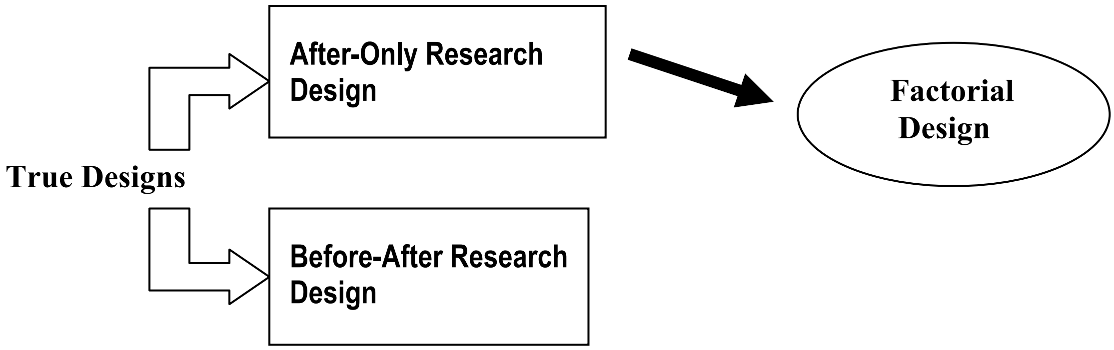

Figure 4.4: Types of true experiments

4.3.1 After-Only Research Design

The After-Only Research Design gets its name from the fact that the dependent variable is measured only once after the experimental treatment. In other words, the posttest is administered once to the experimental group and the control group (see Figure 4.5). It shows an experiment in which the researcher is attempting to show the effectiveness of the inductive method in improving the science problem skills of 17 year old secondary school students. The sample was drawn from a population and randomly assigned to the experimental and control group. The experimental group were taught science using the inductive approach while students in the control group were not taught using the inductive approach. Instead students in this group were taught the same science content using the traditional didactic approach ('chalk and talk' method).

		Treatment (inductive Method)	Posttest (Science Problem Solving Test Scores)
EXPERIMENTAL GROUP	R	X	O
CONTROL GROUP	R	-	O

Note: R – random assignment

Figure 4.5: After-only Research Design

SELF-CHECK 4.2

1. What is the main strength of 'true' experiments?
2. What is the major difference between the two types of true experiments discussed?

4.3.2 Factorial Research Design

The factorial design is an after-design research design that allows the study of two or more independent variables simultaneously and their interactive effects on the dependent variable. To understand the factorial design, a hypothetical example is shown in Figure 4.6. The experiment (2 x 2 factorial design) aims to examine the effectiveness of two teaching methods (Independent Variable A) on performance in history (Dependent Variable) among a sample of 17 year old students of different ability levels (Independent Variable B).

- Method (independent variable A) is made of two methods:
  - Deductive teaching method: In this method, students are presented with a concept followed by the examples.
  - Inductive teaching method: In this method, students are presented with examples and from these examples they derive the concept.
- Ability (independent variable B) is divided into two levels:
  - High ability: based on their academic performance scores.
  - Low ability: based on their academic performance scores.

			Independent Variable A METHODS	
			Inductive (A_1)	Deductive (A_2)
Independent Variable B	**ACADEMIC ABILITY**	High (B_1)	$A_1 B_1$	$A_2 B_2$
		Low (B_2)	$A_1 B_2$	$A_2 B_2$

Figure 4.6: A 2 x 2 Factorial Design

So there are four possible combinations of the two independent variables. Each of these treatment combinations are referred to as *cells* (i.e. A1B1; A2B1; A2B1 and A2B2). Subjects are randomly assigned to these four cells within the design. For the experiment using this factorial design, you are looking for THREE different kinds of effect: the main effect of method, the main effect for ability and the interaction between method and ability.

In this example, you are able to test three null hypotheses:

1. There is no significant difference between the inductive method and deductive method on performance in history (MAIN EFFECT – Method).
2. There is no significant difference between high and low ability students on performance in history (MAIN EFFECT – Ability).
3. There is no significant interaction between method and ability (INTERACTION EFFECT).

			Independent Variable A METHODS		
			Inductive (A_1)	Deductive (A_2)	**MEAN**
Independent Variable B	**ACADEMIC ABILITY**	High (B1)	60.0	50.0	**55.0**
		Low (B2)	40.0	30.0	**35.0**
		MEAN	**50.0**	**40.0**	

Main effects for Method (arrows to 50.0 and 40.0) **Main effects for Ability** (arrows to 55.0 and 35.0)

Figure 4.7: Factorial design showing means for ability and method

The results of the hypothetical experiment are shown in Figure 4.7. The main effects for methods (variable A) showed that there was a significant difference in history performance between students taught the inductive (Mean = 50.0) and the deductive approach (M = 40.0). This means that method had an 'effect' on history performance. There was also main effect for ability (variable B) where a significant difference was observed between high (M = 55.0) and low ability (M = 35.0) students on performance in history. Similarly, it means that ability had an 'effect' on history performance.

However, the interaction effect was not statistically significant. What is an interaction? A psychologist is asked, does listening to a motivation talk improve academic performance? When the psychologist replies, "Yes, but it depends on ……." or "It is more complicated than that," he or she is referring to 'interaction'. An interaction effect tells us about the influence of one independent variable on another. In the case of our hypothetical example, it is; whether the combination of 'method' and 'ability' produced an effect on performance in history. Though the two variables (method and ability) by themselves had a significant effect, the combination of method and ability did not produce an effect on performance in history.

In a 2 X 2 experiment (the hypothetical experiment), you can obtain 8 basic patterns of results (see Figure 4.8):

1. A main effect for *method,* no main effect for *ability* and No *interaction*.
2. No main effect for *method*, a main effect for *ability* and No *interaction*.
3. A main effect for *method*, a main affect for *ability* and No *interaction*.
4. A main effect for *method*, a main effect for *ability* and an *interaction*.
5. No main effect for either *method* or *ability*, but an *interaction*.
6. A main effect for *method*, no main effect for *ability*, and an *interaction*.
7. No main effect for *method*, a main effect for *ability* and an *interaction*.
8. No main effects (*method* & *ability*) or *interaction*.

Figure 4.8: Patterns of effects in a 2 X 2 factorial design

(a) **Main Effect for Method and Main Effect for Ability and No Interaction (No: 3)**
Let us examine what this means with our hypothetical example. The data in Figure 4.7 indicates that you have main effects for both method and ability. Look at the first row. You can see that high ability learners treated with the inductive method (M = 60.0) scored higher than high ability learners treated with the deductive method (M = 50.0). Looking at the next row, you see that low ability learners treated with the inductive method (M = 40.0) scored higher than low ability learners treated with the deductive method (30.0). You can see this in Figure 4.8.

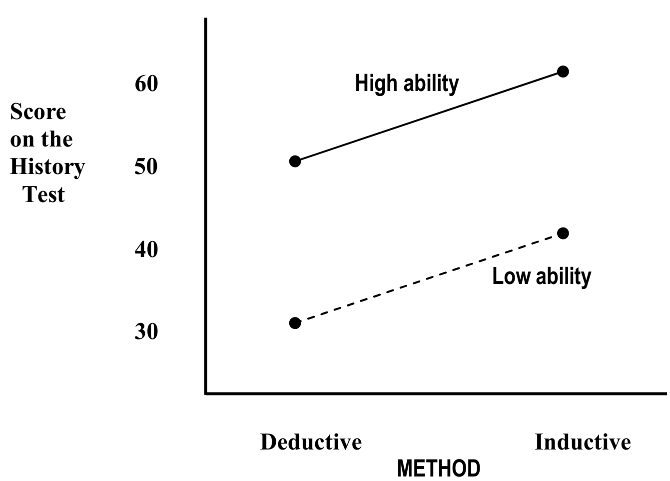

Figure 4.9: Graph showing no interaction

Looking at the columns tell you about the effect of ability. You see that high ability learners treated with the inductive method scored 60.0 compared to their low ability counterparts who scored 40.0 (i.e. 20 more) in the same treatment group. Looking at the second column, you learn that high ability learners (M = 50.0) treated with the deductive method scored higher than low ability learners (M = 30.0) treated with same method. Thus, it appears that in addition to the method main effect, you have an ability main effect.

Finally, you also know that there is no interaction because the effect of method is unaffected by the ability level of students. As Figure 4.7 demonstrates, the effect of method is independent of ability level and the effect of ability level is independent of method of instruction. If you graph the means, your graph should look something like Figure 4.8. The graph confirms what you saw in Figure 4.7. The high ability line is above the low ability line. Similarly, ability increases as shown by the fact that both lines slope upwards as they go from deductive to inductive method. Finally, the graph tells you that there is no interaction between method and ability on performance in history because the lines are parallel.

(b) **No Main Effect for Method and No Main Effect for Ability but an Interaction (No: 7)**

Let us examine what does this mean with our hypothetical experiment. According to Figure 4.10, the means obtained on history performance according to method reveals no significant difference between the inductive method (M=55.0) and the deductive method (M=55.0). Similarly, for ability there was no significance difference between high ability students and low ability students. However, there was an interaction and the interaction was significant (see Figure 4.11). In this figure, you notice that the lines are not parallel (as in Figure 4.9). Therefore, you have an interaction. What is the meaning of this interaction since there was no effect for either method or ability? You would say that method has an effect, but its effect depends on ability level. Alternatively, you could say that ability has an effect but that effect depends on the type of method students had been treated with.

SELF-CHECK 4.3

1. What is the main advantage of using the factorial design?
2. Why the factorial design is considered a true experiment?
3. Identify the difference between main effects and interaction effect?

			Independent Variable A METHODS		
			Inductive (A_1)	Deductive (A_2)	**MEAN**
Independent Variable B	**ACADEMIC ABILITY**	High (B1)	60.0	50.0	**55.0**
		Low (B2)	50.0	60.0	**55.0**
		MEAN	**55.0**	**55.0**	

No Main effects for Method **No Main effects for Ability**

Figure 4.10: Factorial design showing means for methods and ability

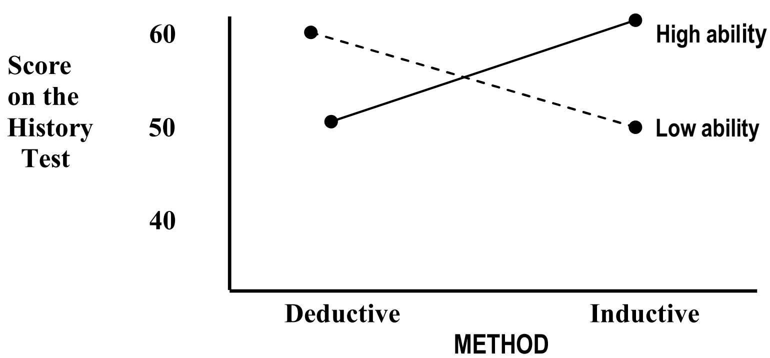

Figure 4.11: Graph showing an interaction

ACTIVITY 4.3

A lecturer doing a experiment finds that students who are given lecture notes but do not attend the lecture, perform better than those who attend the lecture. Refine the study by using a 2 X 2 factorial design.

4.3.3 Before-After Research Design

The Before-After Research Design is perhaps the best example of a true research design that incorporates both an experimental and control group to which the subjects are randomly assigned (see Figure 4.12). This research design is a good experimental design because it does a good job of controlling for extraneous factors such as history, maturation, instrumentation, selection bias and regression to the mean. How is this done? Any history events (e.g. certain events subjects may have been exposed to) that may have produced a difference in the experimental group would also have produced a difference in the control group. Here it is assumed that subjects in both groups have experienced the same set events.

		Pretest	Treatment	Posttest
Experimental Group	**R**	**O**	**X**	**O**
Control Group	**R**	**O**	**-**	**O**

Figure 4.12: Before-after research design

4.4 QUASI-EXPERIMENTAL DESIGN

So far we have examined both weak and strong experimental research designs. However, in educational research there are times when investigators are faced with situation in which all the requirements of a true experiment cannot be met.

For example, sometimes it is not possible to assign students to groups which are a requirement of strong experimental research. Due to logistical reasons it is difficult to randomly assign subjects to groups and so intact groups such as a class may have to be used. Is it still possible to do an experiment despite these limitations? The answer is 'yes'; you can use a quasi-experimental design.

According to Christensen and Johnson (2000), a quasi-experimental design is "an experimental research design that does not provide for full control of potential confounding variables. In most instances, the primary reason that full control is not achieved is because participants cannot be randomly assigned" (p. 255).

4.4.1 Non-Equivalent Control-Group Design

	Pretest	Treatment	Posttest
Experimental Group	O	X	O
Control Group	O	-	O

Figure 4.13: Non-equivalent control-group design

The Non-equivalent control-group design contains an experimental and control group, but the subjects are not randomly assigned to groups (see Figure 4.13). The fact there is no random assignment means that subjects in the experimental group and control group may not be equivalent on all variables. For example, you could have more low ability students in the control group compared to the experimental group. Hence, it may be difficult to establish whether the better performance of the experimental group is due to the treatment or because there are more high ability students in the group.

In the Non-equivalent control-group design both groups are given first a pretest and then a posttest [after the treatment is given to the experimental group]. The pretest score and the posttest score are compared to determine if there are significant differences.

When you cannot randomly assign subjects, you can be sure that extraneous variables or factors will creep into the experiment and threaten its internal validity. [We have discussed in Chapter 3, the factors that threaten the internal validity of experiment]. Do you leave it alone or do something about it?

Knowing that extraneous factors will creep into a quasi-experiment, a good researcher will take steps to ensure that the subjects in the experimental group and control group are as similar as possible, especially on important variables such as academic ability, attitude, interest, socioeconomic status and so forth. How do you about doing this?

Cook and Campbell (1979) propose the following steps to enhance the internal validity of the non-equivalent control-group design or quasi-experiments in general:

- Selection: ensure that subjects in the experimental and control are matched in terms of important variables that may affect the results of the experiment. For example; match subjects in terms of academic ability, IQ, attitudes, interests, gender, socioeconomic background and so forth)
- Testing: ensure that the time period between the pretest and posttest is not too short such that subjects are able to remember the questions given to them earlier.
- History: ensure that events outside the experiment do not affect the experiment. The problem is most serious when only subjects from one of the groups are exposed to such events (e.g. motivation talks, private tuition)
- Instrumentation: ensure that the pretest and the posttest are similar. If a different test is used, you should make sure that the two tests are equivalent in terms of what it is measuring (i.e. high reliability and validity).

4.4.2 Interrupted Time Series Design

The interrupted time-series design requires the researcher to take a series of measurements both before and after the treatment. A single group of subjects are pretested a number of times during the baseline phase, exposed to the treatment, and then posted a number of times after the treatment. 'Baseline' refers to the testing done before the treatment designed to alter behaviour.

A hypothetical example may illustrate how the interrupted time series design is used. Say that you want to determine whether positive reinforcement encourages inattentive low ability learners to be more attentive. You identify a group of 11 year old low ability learners and get them to attend an experimental classroom for at least one period each school day (see Figure 4.14). In this classroom, subjects are taught reading skills in a positive environment where they were praised and rewarded for attentive behaviour that is focussed on the given task activities. Before the students were sent to the positive treatment classroom their behaviour was observed over three sessions in their regular classroom with regards to their attentiveness. This was to obtain baseline data where their behaviour was recorded in its freely occurring state. The treatment lasted for three weeks and after the treatment, subjects were observed for their attentiveness and focused behaviour.

Multiple **PRETESTS**	**TREATMENT**	Multiple **POSTTESTS**
O_1 O_2 O_3	**X**	O_4 O_5 O_6

Figure 4.14: Interrupted time-series design

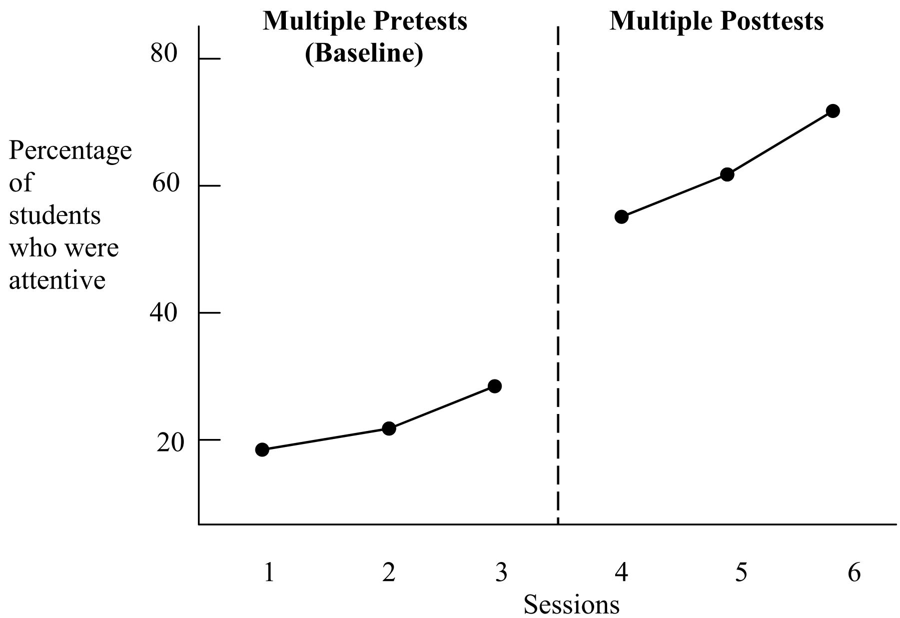

Figure 4.15: Percentage of students observed to be attentive and focussed

The result of the hypothetical experiment is shown in Figure 4.15 which illustrates the percentage of students who were attentive and focussed on the given task. From this graph you can see that the percentage of students who were attentive and focussed who were assessed multiple times prior to and after implementation of the positive classroom environment, making it an interrupted time-series design. This assessment reveals that the percentage of students who were attentive and focussed remained rather constant during the first 3 baseline class sessions, or the class sessions prior to the implementation of the positive classroom environment. After implementation of the positive classroom environment, the percentage of attentive behaviour consistently increased over the next three class sessions, suggesting that the implementation of the positive approach had a beneficial effect on the behaviour of inattentive students.

SELF-CHECK 4.4

1. What is the meaning of non-equivalent in the non-equivalent control group design?
2. How can you enhance the internal validity of quasi-experimental research designs?
3. When would you use the interrupted time-series design?

4.5 ETHICS IN EXPERIMENTAL RESEARCH

During World War II, Nazi scientists conducted some gross experiments such as immersing people in ice water to determine how long it would take them to freeze to death. They also injected prisoners with newly developed drugs to determine their effectiveness and many died in the process. However, these experiments were conducted by individuals living in a demented society and universally condemned as being unethical and inhumane. Research in education involves human as subjects: students, teachers, school administrators, parents and so forth. These individuals have certain rights, such as the right to privacy that may be violated if you are to attempt to arrive at answers to many significant questions. Obviously, this becomes a dilemma for the researcher as to whether to conduct the experiment and violate the rights of subjects, or abandon the study. Surely, you have heard people say: "I guess we are the guinea pigs in this study!". "We are your while rats!".

Any researcher conducting an experiment must ensure that the dignity and welfare of the subjects are maintained. The American Psychological Association published the *Ethical Principles in the Conduct of Research with Human Participants* in 1982. The document listed the following principles:

- In planning a study, the researcher must take responsibility to ensure that the study respects human values and protect the rights of human subjects.
- The researcher should determine the degree of risk imposed on subjects by the study (e.g. stress on subjects, subjects required to take drugs).
- The principal researcher is responsible for the ethical conduct of the study and be responsible for assistants or other researchers involved.
- The researcher should make it clear to the subjects before they participate in the study regarding their obligations and responsibilities. The researcher should inform subjects of all aspects of the research that might influence their decision to participate.
- If the researcher cannot tell everything about the experiment because it is too technical or it will affect the study, then the researcher must inform subjects after the experiment.
- The researcher should respect the individual's freedom to decline to participate in or withdraw from the experiment at any time.
- The researcher should protect subjects from physical and mental discomfort, harm, and danger that may arise from the experiment. If there are risks involved, the researcher must inform the subjects of that fact.

- Information obtained from the subjects in the experiment is confidential unless otherwise agreed upon. Data should be reported as group performance and not individual performance.

SELF-CHECK 4.5

What are some ethical principles proposed by the American Psychological Association with regards to doing experiments involving human subjects?

SUMMARY

- A research design is a plan or strategy specifying the procedure in seeking an answer to the research question.
- 'Weak' research designs do not allow for controlling of extraneous that might creep into the experiment.
- Examples of weak designs: one shot design, one-group pretest-posttest design and non-equivalent posttest-only design.
- 'True' experimental designs enable the researcher to maintain control over the situation in terms of assignment of subjects to groups.
- Examples of true designs: after-only research design, factorial design and before-after research design.
- A quasi-experimental design is a design that does not provide for full control of potential confounding variables.
- Examples of quasi-experimental designs: non-equivalent control-group design and interrupted time-series.
- Researchers conducting experiments involving human subjects should respect the privacy and confidentiality of subjects.

KEY TERMS

Experimental Design

Quasi-Experimental Design
- non-equivalent design
- time series design

True research designs
- after-only design
- factorial design
- before-after design

Weak research designs
- one-shot design
- one-group pretest-posttest
- non-equivalent posttest only

DISCUSSION

1. Make a case for the superiority of true experimental designs.
2. What are quasi-experimental research designs and how do they differ from true experiments?
3. Discuss the circumstances in which researchers have to use intact groups.
4. What can a researcher do to increase the equivalence of subjects in the control and experimental groups in a quasi-experiment design?
5. Graph the following data from an experiment on the effect of lighting and music on anxiety. The scores are means on an anxiety test.

		Music	
		Classical	Rock
Lighting Level	Dim	45	11
	Bright	12	44

Is there an interaction? How do you know?

READINGS

Books

Christensen, L. (1988). E*xperimental methodology.* Boston: Allyn and Bacon Inc. *Chapter 8: Experimental research designs.*

Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches.* Needham Heights, MA: Pearson Education. *Chapter 9: Quasi-experimental and single-case designs.*

Mitchell, M., & Jolley, J. (1988). *Research design explained.* New York: Holt, Rinehart & Winston. *Chapter 7: Expanding the simple experiment: factorial designs.*

Internet Resources

Abrahams, D. (n. d.). *True experimental designs and their meaning.* Retrieved from: http://www.socialresearchmethods.net/tutorial/Abrahams/true.htm

Abrahams, D. (n. d.). *Pre-experimental designs and their meaning.* Retrieved from: http://www.socialresearchmethods.net/tutorial/Abrahams/preex.htm

Connor, T. (n. d.). *Experimental and quasi-experimental research design.* Retrieved from: http://faculty.ncwc.edu/toconnor/308/308lect06.htm