Topic 3 ▶ Experimental Methodology

LEARNING OUTCOMES

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

1. Define what is an experiment;
2. Explain the components of an experiment in education;
3. Identify the threats to internal validity of experiments;
4. Explain how to control for extraneous variables that affect the internal validity of experiments;
5. Describe how random assignment is performed;

▶ INTRODUCTION

"Remember more and think faster with BE SMART"

"Rewarding pre-schoolers with chocolates has improved attention in class"

"The Mental Awareness Approach has proven to be an effective way to help smokers give up the habit"

"Cognitive therapy is an effective method for treating drug addicts"

"Enhancing self-esteem improves academic performance"

Source: faculty.washington.edu/chudler

You may have come across these statements or somewhat similar statements? Note that each statement is making a claim that their proposed method, product, technique or procedure is effective in enhancing human performance. Obviously, you would like to know how they went about proving 'effectiveness'. How does one prove effectiveness? Of all available research methods the experimental method is the best. You may have conducted science experiments in the laboratory or in the field! The experimental method was originally used in the field of agriculture where experiments were conducted to test the effectiveness of various kinds of treatments such as fertilisers, water and sunlight on plant growth. The method is used in medical sciences especially in testing the effectiveness of various kinds of drugs, procedures and therapy on patients. The experimental method is widely used in education in which researchers observe the occurrence of a phenomenon as a consequence of a particular action or intervention.

3.1 THE EXPERIMENTAL METHOD

An experiment is a research method used to determine the effectiveness of a particular action or treatment on a single or group of organisms. To show that a particular **treatment** has an effect or brings about a particular change, the researcher has to control all other factors that might influence the occurrence of the particular change. The experimental method is the best method to show **effectiveness** of a particular treatment (e.g. teaching method, curriculum innovation). Experiments are ideally suited for the task of **causal analysis** (claim to show "cause and effect"). No other method of scientific inquiry permits the researcher to say with confidence that "X (praising young learners) caused Y (to repeat the task) to happen".

Hence, it is important that you use the word "effectiveness" carefully, as it only applies if you are using the experimental method.

See Figure 3.1 which shows a simple experiment to determine whether teaching young learners using analogies (e.g. blood circulation is like a river and its tributaries) "causes" them to perform better academically in science ("effect"). The experiment involves administering a treatment (Independent Variable) such as teaching science using analogies. A pretest (Dependent Variable) is given before the experiment and the same test or equivalent test is given after the experiment. The differences between pretest scores and posttest scores will determine whether teaching using analogies improves performance in science.

PRETEST (Science test)	TREATMENT (Teaching using analogies)	POSTTEST (Science test)

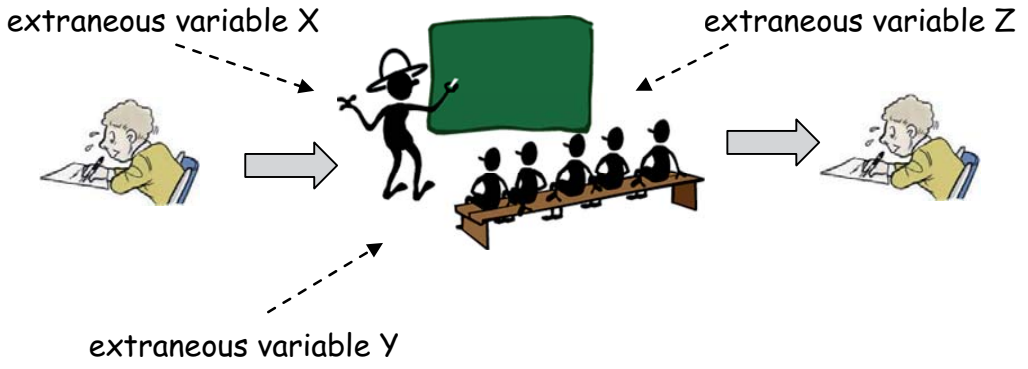

Figure 3.1: A Simple experiment

However, experiments are difficult to conduct. Many experiments in education are concerned with testing the effectiveness of certain interventions or educational practices on student learning, attitudes, perceptions and so forth. A key problem in conducting experiments is establishing suitable **control,** so that any change in behaviour can be attributed only to the treatment introduced by the researcher. Control means ruling out other possible causes for the changes in the behaviour of subjects (see Figure 3.1). There are many **extraneous variables** (irrelevant or unrelated or unconnected factors) that need to be controlled so that they do not contaminate or interfere with the findings of the study. Once an extraneous variable creeps into an experiment, the researcher can no longer draw any conclusion regarding the causal relationship that exists between the independent and the dependent variable (Christensen, 1988).

In education, many experiments are conducted in the classroom (natural setting) and so many factors not related to the treatment may influence performance in the posttest. With reference to Figure 3.1, some students may have discussed with their friends at home concerning the science topic, while others may have viewed a programme on the topic on TV. So, improved performance on the posttest may not be attributed to the treatment but due to the influence of other factors. Therefore, it is necessary to control for the influence of these outside factors or variables in order to attain internal validity.

Some experiments have both an experimental group and a control group. An *experimental group* consists of subjects who are exposed to the treatment. For example, a particular counselling technique is used for a group of juvenile delinquents. The *control group* consists of subjects who do not receive the

treatment (i.e. they are not 'treated' with the counselling technique). Comparison between the experimental group and the control group determines the effectiveness of the counselling technique. In some experiments there may be more than one experimental group; subjects treated with two or three different methods or techniques or procedures are compared with the control group who do not receive any of the treatments. You can also compare the effectiveness of different treatments on the dependent variable.

SELF-CHECK 3.1

1. What is unique about the experimental method compared to other methods of research?
2. What is 'treatment'?
3. What is the difference between an experimental group and a control group? Why do you need these two groups?

3.2 EXTRANEOUS VARIABLES TO BE CONTROLLED TO ENHANCE INTERNAL VALIDITY OF EXPERIMENTS

In conducting experiments, you should ensure that your design has the highest internal validity possible. What is internal validity? The **internal validity** of an experiment is the extent to which extraneous variables (irrelevant variables) have been controlled or ruled out by the researcher. Internal validity is an indication that the results you obtain are caused by the treatment you administered and not some other variable or factor. For example, in your experiment you taught (treated) one group of four year olds with the whole word method of reading and discovered that their reading ability increased by 50% compared to the group who were taught (treated) with the 'phonics method of reading'. How can you be sure that the increase in reading scores of the whole-word method group is DUE to the method taught and not some other factors or variables? Generally, an experiment with high internal validity the probability that the treatment caused the change is higher.

Campbell and Stanley (1963) state that experiments are internally valid when the obtained effect can be attributed to the manipulation of the independent variable. In other words, if the effects (e.g. improved scores in mathematics) obtained in the experiment are due only to the experimental conditions manipulated by the researcher and not to any other variables (factors), the experiment has internal validity. In any experiment there are always some other than the independent variable (treatment) that could influence the observed effects (dependent variable). These variables must be identified and dealt with or held constant. Cook and Campbell (1979) list a number of factors that can threaten the validity of experiments. It is important that you knows these threats so that you can take the necessary steps to control the influence of these outside factors to enhance internal validity.

3.2.1 Time Interval and Threats to Internal Validity

In conducting an experiment, a pretest and posttest is administered to subjects undergoing the treatment. The time interval between the pre- and post measurement of the dependent variable can introduce extraneous factors (see Figure 3.2).

- The first is **History** which includes events that have occurred in the subjects' environment between the pre-test and the posttest that might affect the scores. For example, the subjects may have experienced events during the time lapse that affected their attitude and this is reflected in the scores of the dependent measure. Generally, the longer the duration between the pre- and the posttest, the greater the possibility of history threatening internal validity. But even short time lapses can generate the history effect.
- The second is **Maturation** in which subjects may change between the pre-test and posttest resulting in subjects becoming more mature. The change could be both biological and psychological such as age, learning, fatigue, boredom and hunger that are not related to specific external events but reside within the individual.
- The third is **Instrumentation** whereby change in instruments used in the pretest and the posttest can lead to changes in measurement. For example, an easier test used in the posttest will result in better performance in the posttest because of the instrument and not the treatment. Alternatively, in your pretest you used a multiple-choice test to measure the effect of the treatment. In the posttest an essay test was used.
- The fourth is **Testing** whereby subjects remember the questions in the pretest and if the same test is given as a posttest, the chances are they may score higher in the posttest i.e. they have become "test-wise". The time period

between the pretest and the posttest should not be too short such that subjects can recall.

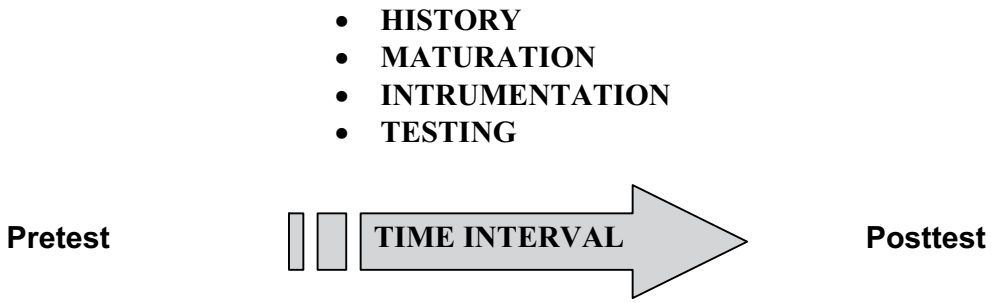

Figure 3.2: Time interval between pretest and posttest and threats to internal validity

3.2.2 Other Threats to Internal Validity

Besides the above four factors threatening internal validity, there are three other factors. First is **Mortality** which is sometimes referred to as "attrition" when subjects drop out from the experiment which can affect the experiment. This is especially serious when subjects of a particular characteristic (e.g. high ability) systematically drop out. Second is **Selection Bias** when the subjects selected for the experimental group and the control group are not equivalent before the treatment leading to a misleading conclusion. For example, if the experimental group consists of 50% high ability subjects while the control group consists of only 25% high ability subjects, higher performance on the posttest may not be attributed to the treatment but due to non-equivalent subjects in terms of ability. Third is **Regression to the Mean** when subjects with extreme scores on a test are selected, there is a likelihood that when they are retested later on a measure that is correlated with the first test, their scores will move towards the mean. For example, students who performed poorly are selected for training; their average posttest scores will be higher than their pretest scores because of statistical regression even if no training were given.

3.2.3 Subject-Experimenter Effects to be controlled

You should know that in an experiment the experimenter or researcher interacts with the subjects. We assume that the subjects taking part in the experiment will listen to the instructions and perform all tasks according the way you planned it. However, you well know that in reality this does not happen because your subjects are of varied backgrounds and have their own perceptions and opinions. This may lead to subjects responding to the experiment in different ways that

may affect the experiment which has been termed as 'subject-experimenter effects'.

(a) **Subject Effect**

The perception of subjects when they enter the experiment can affect how they respond to the tasks required of them. Their perception of the purpose of the experiment, the task required and the rumours they hear about the experiment may cause them to behave differently. For example, subjects who realise that the experiment is about speed of learning and intelligence. There is the tendency to learn the material presented as rapidly as possible to appear intelligent. Similarly, if the task suggests something about emotional stability, subject may respond in such way as to appear most emotionally stable.

If there is an experimental group and control group, the is the tendency for subjects in the experimental group to succumb to the *novelty effect* because the treatment given is different from what they are used. Subjects tend to be enthusiastic especially in the beginning which may wear off as the treatment continues.

(b) **Experimenter Effect**

The experimenter has a motive for conducting the experimenter. He or she is attempting to uncover the laws of human behaviour through experimentation. Towards this goal, the experimenter expects subjects to be perfect respondents who will cooperate and follow instructions carefully. The experimenter may be too keen to obtain findings that confirm the hypotheses and this desire is communicated unconsciously to subjects. The subtle cues presented by the experimenter are picked up by subjects and influence their performance in the direction desired by the experimenter. Certain attributes of the experimenter has shown to influence subjects. For example, in some experiments young children respond more readily to women experimenters compared to their male counterparts.

If there is an experimental and control group, the researcher if not careful may pay special attention to subjects in the experimental group which may influence their behaviour. This is called the *Hawthorne Effect.*

SELF-CHECK 3.2

1. What is meant by internal validity and why is it an important ingredient in experimental research?
2. Identify the major extraneous variables that need to be controlled with an experiment?
3. How do these extraneous variables affect the internal validity of experiments?
4. Explain how subjects and the experimenter can bias the results of an experiment?

3.3 RANDOM ASSIGNMENT TO ENHANCE INTERNAL VALIDITY

An important issue when conducting experiments is how subjects are assigned to the groups. This is important because it determines whether your study is a true experiment or a quasi-experiment. [*We will discuss this issue in Chapter 4: Experimental Research Designs*]. Random assignment means that each sampling unit (e.g. student, teacher, class, etc) has an equal chance of being selected in the experiment. In designing an experiment, you should ensure random assignment as it is the best technique available in establishing that the two or more groups are equivalent. Equivalent means that the subjects in the two or more groups have more or less similar characteristics, such as similar ability levels, similar attitudes, similar number of males and females, similar experiences, similar socio-economic backgrounds and so forth. If the subjects are not randomly assigned, there is the possibility that you may have disproportionately high ability subjects in one group. If they score high after the treatment, it may be attributed to the larger number of high ability subject rather than the treatment.

A popular technique used to ensure random assignment is to use the Table of Random Numbers. Say you have 70 subjects to be assigned to two groups (see Figure 3.3). Assign number 1 to 70 to the subjects. Then, refer to the Table of Random Number (see Table 3.1) and select a starting point, let's say you take the third column which has the numbers 26, 54, 37, 98, 39 and so forth. You will select subject no. 26 assigned to Group 1 followed by subject no. 54 assigned to the Group 2. Of course you will ignore number 98 because it is outside the 70 subjects. You will continue this procedure until all 70 subjects have been assigned to the two groups.

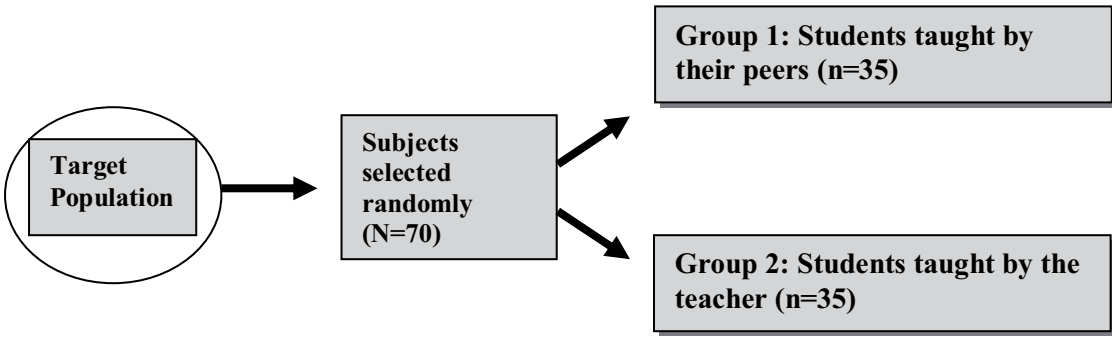

Figure 3.3: Random assignment of subjects to two groups

Table 3.1: Table of Random Numbers

23	34	26	91	73	93	83	59	50	51
76	79	54	45	65	13	11	56	91	27
68	57	37	38	45	45	04	85	66	12
45	25	98	63	52	23	03	36	06	08
89	3	39	34	91	94	12	39	13	31
90	26	83	26	21	34	82	07	34	67
23	61	64	65	37	06	54	26	29	75
87	82	51	02	95	64	62	35	96	49
90	71	25	86	62	39	53	49	48	52
12	38	67	09	67	31	45	40	28	31

SELF-CHECK 3.3

1. Why is it important that subjects are assigned randomly in an experiment?

3.4 OTHER TECHIQUES TO ENSURE GROUPS ARE EQUIVALENT

One of the difficult tasks for a researcher using the experimental method is getting two or more equivalent groups. Imagine the difficulty of finding two people who are similar on every characteristic such as IQ, attitude, aptitude, mathematical ability and so forth. As mentioned earlier, random assignment is a powerful way to ensure that subjects assigned to the various groups will have more or less similar characteristics. There are other techniques used to increase the probability of subjects in two or more groups are equivalent.

3.4.1 Matching

Determine a particular factor, for example, academic performance which is measurable and categorised as High and Low. From the sample, select two High Ability subjects and randomly assign them to the control group and the experimental group. Next, select two Low Ability subjects and assign them randomly to the control group and the experimental. Continue doing this until all subjects have been assigned and your two groups are matched in terms of academic performance.

Another technique of matching is to give the pretest and based on the scores obtained assign subjects to the control group and the experimental group. However, you should ensure that the average score or mean score of the pretest should be the same for the two groups. e.g. two subjects with mean of 23; two subjects with mean 30; two subjects with mean 34 and so forth.

3.4.2 Holding One or More Variables Constant

Another method is to hold a particular variable constant. For example, in an experiment you have difficulty ensuring that the control group and the experimental have an equal number of high socioeconomic and low socioeconomic subjects. You could take only low socioeconomic subjects and assign them randomly to the control group and experimental group if you are not interested in comparing high and low socioeconomic subjects (see Figure 3.4). What you have done is to eliminate the socioeconomic factor or variable by including only low socioeconomic subjects; i.e. **controlling by holding** socioeconomic constant across the two groups you are comparing.

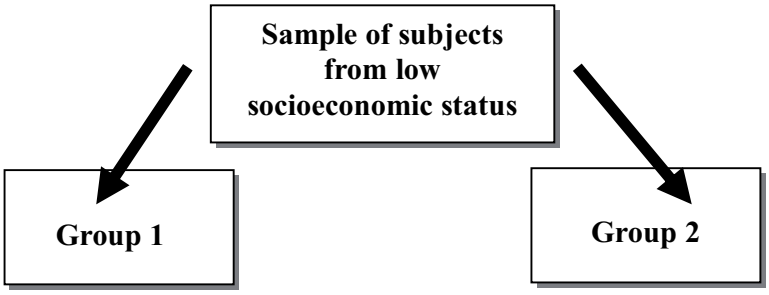

Figure 3.4: Control by holding a variable constant

3.4.3 Including an Extraneous Variable in the Research Design

You could control a variable by including it in the design and making it another independent variable. For example, you design an experiment to test the effectiveness of getting students to define concepts using their own words on performance in economics (see Figure 3.5). However, you find it difficult to control for prior knowledge in economics among your subjects. You could include only those who have low prior knowledge based on a test on economics you administered or you could categorise prior knowledge as High, Medium and Low based on test scores and treat prior knowledge levels as an independent variable. However, you should use this technique only if you are interested in the influence of prior knowledge on performance. What you have done is to control the influence of prior knowledge on other independent variables by including it in the research design.

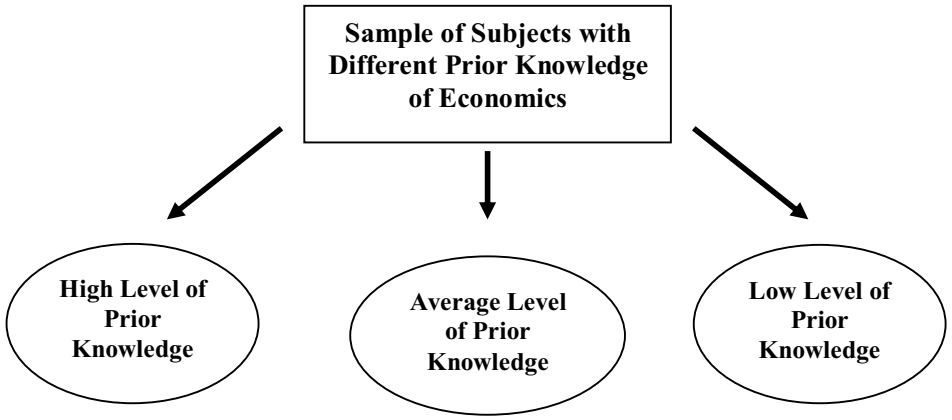

Figure 3.5: Including extraneous variable in the design

3.4.4 Analysis of Covariance (ANCOVA)

It is a statistical method used to ensure that the subjects in the control group and the experimental group are equivalent on various factors. ANCOVA adjusts the scores on the dependent measure for the differences found on the pretest and statistically equates the subjects in the control and experimental group. For example, you are conducting a study on the effectiveness of metacognitive training on the critical thinking skills of Form 4 students. However, you find that some subjects in your experiment are high achievers while the others are low achievers which may influence performance on the critical thinking test (i.e. the dependent variable). To ensure that all subjects in the control group and the experimental group are equal in academic achievement, the ANCOVA is used to

adjust scores on the critical thinking test for the difference in academic achievement..

"You should keep in mind that ANCOVA is an imperfect statistical technique for equating experimental groups prior to the treatment period. Only the variables that are measured can be used as covariates. The groups may differ on other variables, but if these variables have not been measured, they cannot be entered into the ANCOVA" (Borg, W., & Borg, M. (1988). *Educational research: An introduction, p.684*).

SELF-CHECK 3.4

1. Besides randomisation, what are three techniques of increasing the probability of subjects two or more groups are equivalent?
2. Explain the differences between these three techniques.

3.5 HYPOTHESIS TESTING

A hypothesis is a statement created by a researcher to speculate the outcome of the research that he intends to carry out. In other words, a research hypothesis is a conjecture about the presumed relations between the variables under study. A research usually begins with the research problem. The research problem may be framed in the form of a research question. However, the research question may be too broad or not specific enough for the purpose of conducting statistical testing. The conversion of a research question into the form of a hypothesis makes it more realistic and testable.

There are two types of hypothesis, the hypothesis or alternative hypothesis (denoted by H_1 the null hypothesis (denoted by H_0).

Hypothesis testing involves the following steps:

(a) State the null hypothesis and the alternative hypothesis;

(b) Select a research method that the testing of the null hypothesis to be carried out;

(c) Gather the empirical data;

(d) Use inferential statistical calculation to derive at one of the following possible outcomes:

   (i) Outcome 1: Reject the null hypothesis in favour of the alternative hypothesis; or
   (ii) Outcome 2: Do not reject the null hypothesis

3.5.1 The Research Hypothesis and the Null Hypothesis

The research hypothesis or the alternative hypothesis is the experimental outcome that the researcher conjectures. For example, a researcher believes that a certain treatment (e.g. inductive approach) has a positive effect (e.g. enhancing creative thinking ability) of primary school students. In order to test his conjecture, he carries out an experimental study involving two groups, the experimental group and the control group. The experiment group receives the inductive approach treatment but the control does not undergo such treatment. At the end of the experiment, both groups were evaluated using a test instrument to measure creative learning abilities of both groups.

It is however impossible to test the research hypothesis directly. It is necessary first state a null hypothesis and then to assess the probability that this hull hypothesis is true. Here, the null hypothesis states the negation of what the researcher conjectures. In this example, the null hypothesis states that there is no difference between the two groups in their creative thinking ability (test scores). Statistically, the hypothesis can be expressed as follows:

$H_o: \mu_1 = \mu_2$

OR

$H_o: \mu_1 - \mu_2 = 0$

Where μ_1 is the mean test score for the experimental group (Group 1);
and μ_2 is the mean test score for the control group (Group 2)

Both equation A and equation B indicate that there is no significant difference between the mean test scores of group 1 and group 2.

In this example, since the researcher conjectures that the inductive approach helps in improving creative thinking ability, the alternative hypothesis therefore can be statistically expressed as follows:

$H_o: \mu_1 > \mu_2$

OR

$H_o: \mu_1 - \mu_2 > 0$

3.6 TEST OF SIGNIFICANCE

In order to enable you to reject the null hypothesis, it is necessary to analyse the data **statistically.** Why is this necessary? For example, in your experiment you obtained the following:

	Mean	Standard Deviation
Experimental Group	30.4	3.7
Control Broup	28.3	4.1

To the naive person, he or she might conclude that the experimental group performed better than the control group because the mean score is higher by 2.1 and so that the treatment is effective. This is misleading because it is likely that the differences in the mean between the experimental group and control group could have occurred by chance. In order for you to accept or reject the null hypothesis, it is necessary that you analyse the data statistically because you want to be sure that the treatment administered produced a real effect. How do you determine that the difference between the two groups is caused by the treatment and not some other extraneous variable? You could repeat the experiment and see if you get the same results which will provide evidence of the reliability of the obtained findings.

However, this is not an economical approach and for this reason statistical tests are preferred. The test of significance enables one to determine whether the amount of difference between the two groups is due to chance or due to the treatment. Does a large difference between the mean score of the experimental and control group indicate that the difference is real? Even large differences could occur by chance, although the probability of this happening would be very low. The most common practice is to state a significance level that must be reached; which is a statement of the probability that an observed difference is chance difference. The most common significance levels are .05 and .01; regardless whether you are using the t-test, F-test or the chi-square.

If you decide from the onset of the experiment that the .05 significance level is to be used, it means that you will accept as a real difference only one that is so large that it could have occurred by chance only 5 times in 100 (i.e. 95% not due to chance). If the .01 significance level is selected, then the difference can be expected to occur only 1 time in 100 by chance (i.e. 99% not due to chance).

SELF-CHECK 3.5

1. Why is the statistical test of significance used to determine differences between means?
2. Explain the difference between .05 and .01 level of significance.

ACTIVITY 3.2

Age Group	Pretest Mean	Posttest Mean
Males	52.4	57.2
Females	53.1	64.5*

note: * significant at $p < .05$

The table above shows the pretest and posttest means on a critical thinking skills test. The subjects were taught critical thinking skills one period (40 minutes) a week for six weeks.

1. Give a title for the study.
2. State TWO null hypotheses based on the data above.
3. State TWO conclusions based on the findings.
4. What is the independent variable and the dependent variable?
5. Provide an operational definition for the treatment.

SUMMARY

- An experiment is a research method used to determine the effectiveness of a particular action or treatment on a single or group of organisms.
- The experimental method is the best method to show effectiveness of a particular treatment.
- The internal validity of an experiment is the extent to which extraneous variables have been controlled or ruled out by the researcher.
- History, maturation, testing, selection and instrumentation threaten the internal validity of experiments.
- Random assignment means that each sampling unit has an equal chance of being selected in the experiment.
- Random assignment increases the likelihood that groups are equivalent.
- Other methods of ensuring equivalence of groups are; matching, holding variable constant, including variable in the design and ANCOVA.
- A hypothesis is a conjecture (guess or speculation) about the presumed relations between variables.
- The test of significance enables one to determine whether the amount of difference between two groups is due to chance or due to the treatment.

KEY TERMS

Equivalent Groups
- matching
- holding a variable
- including the variable
- ANACOVA

Experimenter effect

Hypothesis Testing
- null hypothesis
- directional hypothesis

Internal Validity
- selection
- maturation
- testing
- instrumentation
- history

Random Assignment
- Table of random number

Subject effects

Test of Significance

The Experiment
- pretest & posttest
- control & experimental groups

DISCUSSION

1. Explain when you would use an experiment in educational research.
2. What do you mean by the statement that 'Experiments allow the researcher to make causal statements'?
3. Why should you be concerned about the internal validity of an experiment?

READINGS

OUM'Digital Library

Bechhofer, F. (2000). *Principles of research design in the Social Science.* London: Routledge.
Chapter 2: Experiments [available at eBrary].

Books

Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches.* Needham Heights, MA: Pearson Education.
Chapter 8: Experimental research.

Mitchell, M., & Jolley, J. (1988). *Research design explained.* New York: Holt, Rinehart & Winston.
Chapter 4: Internal validity: why researchers value experimental designs.

Internet Resources

Abrahams, D. (n. d.). *Introduction to research design*. Retrieved from: http://www.socialresearchmethods.net/tutorial/Abrahams/sbk16.htm

Abrahams, D. (n. d.). *Design notation*. Retrieved from: http://www.socialresearchmethods.net/tutorial/Abrahams/dnote.htm