Chapter One
Introduction

1.1. Definition of Experimental Psychology

In simple terms, experimental psychology could be defined as the scientific branch of psychology that aims at the scientific investigation of behaviour and mental processes mainly based on experimentation. Alternatively, Levin and Henrichs (1995) defined experimental psychology as a set of various basic methods, procedures, and analytic tools used in the scientific study of behaviour and mental processes to study topics such as perception, memory, learning, cognition, social behaviour, and animal behaviour. However, in spite of the definitions it is important to know that experimental psychology does not solely depend up on experimentation.

- Experimental psychology is the study of psychological issues that uses experimental procedures
- It is an area of psychology that utilizes scientific methods to research the mind and behaviour
- It refers to a specific methodological approach to the study of psychology

From the above definitions, it is important to define the following core concepts:

- *Behaviour:* refers to reactions or actions of people or (sometimes animals) that can be directly observed
- *Mental processes:* internal cognitive processes such as thinking, decision making, judgment, evaluation, etc.
- *Experiment:* systematic manipulation of independent variable over dependent variable in order to observe the effect of this manipulation upon the behaviour studied (DV).

Experiment is mostly used to check cause and effect relationship between variables, finding that certain kind of events are predicted under a certain situation in a regular basis. Experiment requires:

- Manipulation or intervention: Manipulation refers to the act of deliberately varying the level of independent variable.
- requires testable hypothesis or prediction
- Consistent procedure: The treatment, instruments, and all conditions should be similar for all participants in a similar condition.
- careful and systematic observation or recording the desired behaviour
- comparison of treatment conditions
- control: keeping constant all other factors (extraneous variables)to affect behaviour

- objectivity (the results must not be biased based on the expectation of the researcher and other factors)

1.2. Variables in Experimental Psychology

A variable refers to things that can be measured and assume different values. The things that are expected to change are known as variables. There are different types of variables in psychological experiment.

- ***Independent variable***: the variable that is manipulated and whose changes are supposed to produce changes on another variable. The independent variable is a variable which occurs independent of the behaviour of the participants. IV is predetermined by the experimenter. The following are some of the independent variables in psychological experiment:
  - ***Situational variables***: Situational variables refer to different features in the environment that participants might encounter. For example, the presence and absence of others, room size, temperature, mood etc.
  - ***Task variables***: variables that base on the type of tasks performed by subjects (participants). For instance, participants are ordered to do different kinds of activities on memory, intelligence tests, etc. Similarly, animal experiments may involve mazes with different degree of complexity, different types of stimuli in a perception study, and so on.
  - ***Instructional variables*** are manipulated by asking different groups to perform a particular task in different ways. For example, the experimenter may measure the same outcome following different instructions in responding to the independent variable.
  - ***Subject variables*** (*aka* non manipulated variables, natural group variables): refer to already existing characteristics of the individuals participating in the study. For example, sex, age, socioeconomic class, ethnicity, intelligence, physical or psychiatric disorder, and personality.
- ***Dependent variables***: the outcome variable as influenced by the level of independent variable. DV is influenced by participants' behaviour. According to Bordens and Abbot (2005) dependent variables could be classified as follows:
  - *Behavioural measures*: refers to the overt, actual, and recordable actions or behaviours of subjects. For example, measuring the frequency of helpful or aggressive behaviours, frequency of errors, and number of recalled words, etc. Latency

measures (the amount of time needed in doing something, reaction time) such as, how long it takes to help (respond); and number of errors or learning trials are examples of behavioural measures.

- *Physiological measures*: involves the use of special technology products to measure humans (and animals) bodily functioning. For example, we can use position emission tomography (PET) scan, electroencephalogram (EEG), and functional magnetic resonance imaging (fMRI) are some of the commonly used equipment in experimental psychology.
- *Self-report measures*: when participants report their internal feelings, sensations, attitudes, perceptions, etc. Rating scales such as (Likert, and Guttmann) Q-sort questions (when we form qualitative categories and sort questions into these listed categories) are examples of self-report measures.

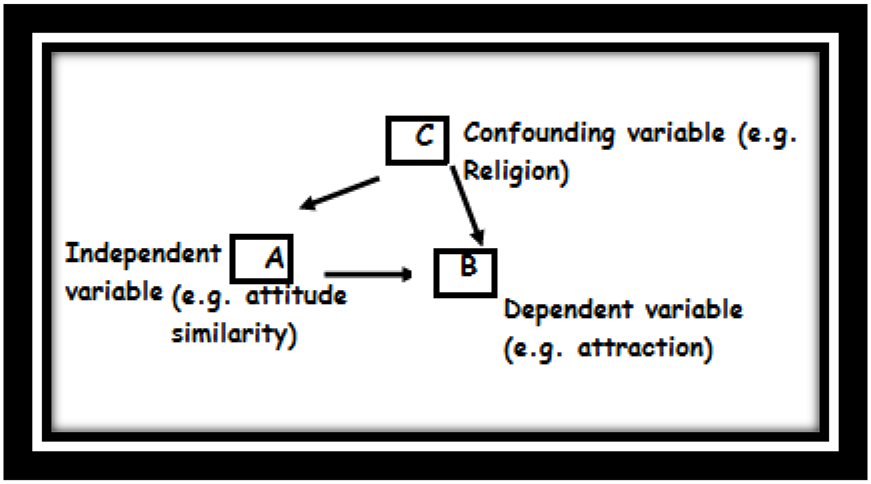
- ***Control variables***

Control variables refer to all extraneous variables kept constant in order for the sake of internal validity. All circumstances other than the independent variable should be kept similar throughout the experiment.

- ***Random variables***

Random variables are variables those could not be totally controlled but expected to vary randomly. For example, mood of participants could be considered as a random variable. Random variation could occur if the participants are randomly assigned. The effect of mood could be zero if there could be equal magnitude of mood change in both control and experimental group.

- ***Confounding variables***: any variable or condition that systematically changes with the independent variable. The effect of confounding variables on the dependent variable could be measured simultaneously with the independent variable.



A researcher may conclude that interpersonal attraction was mainly a result of attitude similarity. However, religion may serve as a confounding variable by systematically varying with the attitude measure.

- ***Extraneous variables***

Extraneous variables are all variables contributing for a change in the dependent variable other than the independent variable. Confounding variables, are examples of extraneous variables.

1.3. History of Experimental Psychology

The historical development of experimental psychology is much related to psychophysics. Psychophysicists are mainly concerned with the study of the relationship between physical stimuli and peoples psychological experience.

- Wilhelm Wundt is best known for the establishment of the first psychological laboratory

Different types of methods are used in research, which loosely fall into 1 of 2 categories.

Experimental (Laboratory, Field & Natural) &
Non experimental (correlations, observations, interviews, questionnaires and case studies).
All the three types of experiments have characteristics in common. They all have:

- an independent variable (I.V.) which is
- a dependent variable (D.V.) which is
- there will be at least two conditions in which participants produce data.

Note – natural and quasi experiments are often used synonymously but are not strictly the same, as with quasi experiments participants cannot be randomly assigned, so rather than there being a condition there is a condition.

Laboratory Experiments

These are conducted under controlled conditions, in which the researcher deliberately changes something (I.V.) to see the effect of this on something else (D.V.).

STRENGTHS

Control – lab experiments have a high degree of control over the environment & other extraneous variables which means that the researcher can accurately assess the effects of the I.V, so it has higher internal validity.

Replicable – due to the researcher's high levels of control, research procedures can be repeated so that the reliability of results can be checked.

LIMITATIONS

Lacks ecological validity – due to the involvement of the researcher in manipulating and controlling variables, findings cannot be easily generalised to other (real life) settings, resulting in poor external validity.

Field Experiments

These are carried out in , in which the researcher something (I.V.) to see the effect of this on something else (D.V.).

STRENGTHS

Validity – field experiments have some degree of control but also are conducted in a natural environment, so can be seen to have reasonable internal and external validity.

LIMITATIONS

Less control than lab experiments and therefore extraneous variables are more likely to distort findings and so internal validity is likely to be lower.

Natural / Quasi Experiments

These are in , in which the researcher measures the effect of something which is to see the effect of this on something else (D.V.). Note that in this case there is no deliberate manipulation of a variable; this already naturally changing, which means the research is merely measuring the effect of something that is already happening.

STRENGTHS

High ecological validity – due to the lack of involvement of the researcher; variables are naturally occurring so findings can be easily generalised to other (real life) settings, resulting in high external validity.

LIMITATIONS

Lack of control – natural experiments have no control over the environment & other extraneous variables which means that the researcher cannot always accurately assess the effects of the I.V, so it has low internal validity.

Not replicable – due to the researcher's lack of control, research procedures cannot be repeated so that the reliability of results cannot be check.

Chapter Two
Reliability and Validity

2.1. Reliability

Reliability is the degree to which an assessment tool produces stable and consistent results. In other words, reliability can be defined as the relative absence of measurement error in a measuring instrument; reliability is then associated with random or chance error. Error and reliability are opposite sides of the same coin. The more error the less stable and less accurate the measurement. Reliability is the accuracy, stability, and relative lack of error in a measuring instrument.

2.1.1. Classical Test Theory

Classical test theory (CTT) states that participants' obtained score in a given test (X) is composed of two components; true score component (T) and error score component (E).

Assumptions of test theory

A. The true score is a measure of a stable and continuing characteristic. It is necessary to assume that the students' performance on the test has not been simply a chance indication of a momentary preference, but rather that is an indication of a stable ability or trait.
B. The error component of a score is a randomly determined variable. Each test score reflects not only a student's true ability, but also some degree of error or incorrect assessment of a student's ability.

According to classical test theory three scores need to be considered for any candidate: the observed score. The true score, and the error score.

Observed score: the total score (including the true and error score) obtained from the test.

The true score: reflects the real attribute that is measured by the test.

Error: refers to measurement errors having two basic forms:

- *Random error:* error caused by varying factors across the sample (e.g. mood). It does not affect mean score.
- *Systematic error*: is caused by systematic factors (causes that can be traced) to affect mean scores consistently (e.g. test problem). This is often known as *bias*.

$$X = T + E$$

$$E = Er + Es$$

Therefore, $X = T + Er + Es$

Where; x= observed/obtained score, *T*= true score, *Er*= random error and *Es*= systematic error

Reliable results are dependable, predictable, accurate, stable and consistent. For this reason, reliability is often used interchangeably with the term dependability, predictability, accuracy, stability and consistency.

Reliability=Consistency

The smaller the error score, the closer the observed score approximates to the true score and the greater the reliability.

Perfect reliability: When the true variance is equal to the observed variance ($S^2T=S^2X$) the ratio has a value of +1.0. It means there is no error variance.

Nil reliability: when there is no true variance present, the ratio has a value of zero. It means the observed variance is entirely error.

Reliability coefficient

The reliability coefficient reflects true score variability. It is symbolized with "r_{xx}".

In contrast to other correlation coefficients, the reliability coefficient is never squared to interpret it but is interpreted directly as a measure of true score variability.

A test's true score variance is not known, however, and reliability must be *estimated* rather than calculated directly. It is assumed that *consistent variability is true score variability, while variability that is inconsistent reflects random error*. For example, reliability coefficient of .84 indicates that 84% of variability is due to true score and 16% is due to measurement error.

A reliability coefficient does not provide any information about what is actually being measured by a test. Whether the test is actually assessing what it was designed to measure is addressed by an analysis of the test's ***validity***.

2.1.2. Methods for Estimating Reliability

The selection of a method for estimating reliability depends on the nature of the test.

1. Test-Retest Reliability

The test-retest method involves measuring the same group of examinees on two different occasions and then correlating the two sets of scores. Its reliability coefficient indicates the coefficient of stability (the degree of stability over time).

The correlation coefficient for two instruments is measured by using

Look at the following correlation coefficient for score of seven students from test 1 and test 2.

S.No.	Score from *Test 1* (X)	Score from *Test 2* (Y)	X^2	Y^2	XY
1	**5**	**6**	25	36	30
2	1	3	1	9	3
3	4	3	16	9	12
4	5	5	25	25	25
5	6	7	36	49	42
6	3	3	9	9	9
7	4	4	16	16	16
	ΣX=28	**ΣY=31**	**ΣX^2=128**	**ΣY^2=153**	**137**

$$r = \frac{\Sigma XY - \Sigma X(\Sigma Y)/n}{\sqrt{[\Sigma X^2 - \frac{(\Sigma X)^2}{n}][\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}]}}$$

$$r = \frac{137 - 28(31)/7}{\sqrt{[128 - \frac{(28)^2}{7}][153 - \frac{(31)^2}{7}]}}$$

$$r = \frac{137 - \frac{28(31)}{7} = 137 - 124 = 13}{\sqrt{\left[128 - \frac{(28)^2}{7}\right]\left[153 - \frac{(31)^2}{7}\right] = 16(15.71) = 251}} = \frac{137 - 124}{\sqrt{16(15.71)}}$$

$$r = \frac{13}{\sqrt{251}} = \frac{13}{15.84} = \mathbf{0.82}$$

Sources of Measurement error for test-retest reliability

The major threats of test-retest reliability are:

- Time gap between the two tests
- Random factors (e.g. Changes in mood, motivation, etc.)
- Carry-over effects: remembering answers from the first tool.
- Practice effect: practice can increase the score from the second test.
- Attrition: subjects may leave in the retest condition.

Test-retest reliability is appropriate to measure attributes that are relatively stable over time (e.g. aptitude) rather than attributes affected by repeated measurements (e.g. mood, creativity, motivation, etc.)

2. *Alternate (Equivalent, Parallel) Forms Reliability*

In alternate forms reliability, we use two equivalent forms of the test administered to the same group of testes and then, the two sets of scores are correlated. This is used to control the effect of carry-over effect.

Form A =Form B

The alternate forms reliability is termed as *coefficient of equivalence* (when the tests are given at about the same time) unless it is termed as *coefficient of stability*.

Source of measurement error for alternate forms reliability

- Random factors
- Specific knowledge from specific version of the test

3. *Delayed Alternate (Equivalent, Parallel) Forms Reliability*

It is similar to the alternate or equivalent forms of reliability but in delayed alternate form, different versions of the test are administered on two different occasions.

Here, we measure the *coefficient of stability (*repeatability of a test over time) due to time gap in administration of the two forms.

Factors affecting test-retest also tend to affect delayed equivalent reliability due to repeated measurement.

Forms of tests in pre-test and post-test

The possible order of the presentation of the test can be as follows:

$Form\ 1 \quad Time\ 1 \quad (Group\ A)$

$Form\ 2 \quad Time\ 2 \quad (Group\ B)$

OR

A111	A112
B111	B112

Sources of Error in Equivalent Reliability Estimation

- Specific events
- Differential knowledge of testees (some may know more items from Form 1 as compared to others)
- Difficulty in developing two version that are really equivalent.

4. *Inter-rater (observer) reliability*

Inter-rater reliability is a method of comparing observations (ratings) of two or more independent observers. This is used often for highly subjective tests (e.g. body movement, emotion, projective tests, etc.). Here, we calculate the correlation coefficient (*kappa coefficient*), *coefficient of concordance* or percentage of agreement. For example, consider ratings for check lists of a behavior among Observer 1 and 2:

	Level (Rating)		
Behaviour	High	Moderate	Low
Jumping	O1	O2	
Laughing		O1, O2	
Moving			O1, O2
Sleeping	O1, O2		
smiling		01,02	

Now, let us calculate the percentage of agreement:

$$r = NA/NT * 100 = 4/5 = 0.8 * 100 = 80$$

Where: NA: no of times the observers agree

NT: number of possibilities to agree

Threats to inter-rater reliability

- Number of response categories (as categories increase, the consensus decreases)
- *Consensual observer drift*, when the raters influence each other's ratings so that they rate the behaviour consistently in a similar way.
- *Observer drift:* when an observer(s) rate a behaviour in consistently different way from other raters.

5. Internal Consistency Reliability

Internal consistency reliabilities use single measurement test administered to the sample at the same time by estimating how items that reflect the same construct show similar result. It is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.

Both parallel and test-retest reliability measures are time consuming that is why we use only the same instrument. There are different internal consistency estimates. The following are the major:

A. Split-half reliability

Split half reliability measures consistency within the test by dividing the test into two equal halves. Then, correlation is calculated between the halves.

Split-half requires large sample of questions.

The items can be divided by using the following methods:

- ✓ Mid-point method
- ✓ Odd-even method, the most common method
- ✓ Random method

Look at the following for creating split-halves based on odd-even method.

X_1	X_2
Q1	Q2
Q3	Q4
Q5	Q6
Q7	Q8
Q9	Q10

We can calculate split-half by using the following three formulas:

a. *Spearman-Brown prophecy formula*
b. *KR-20*
c. *KR-21*

B. **Average inter-item correlation**

It is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each *pair* of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.

2.2. Validity

2.2.1. Definition of validity

Validity refers to the degree to which particular instrument in fact measures what it intends to measure. Validity could be secured through triangulation, selecting appropriate sampling methods, and relevant statistical analysis.

- Validity refers to **the principle of studying what we really want to study**.
- The approximate truth of propositions, assumptions, and inferences.
- Simply validity refers to the process of confirming that we are measuring what we want to measure.

In qualitative research validity could be achieved through honesty (not fabricating data), in-depth information, triangulation, etc.

There are different kinds of validity. Internal validity, content validity, predictive validity, construct, validity, etc are few of them.

2.2.2. Internal Validity

Internal validity refers to the validity related to the experiment. Internal validity is related to the internal consistency in the study. This is to mean that validity indicates a change in the dependent variable is caused only by the change of independent variable. If extraneous or confounding variable are attributed to the results of the dependent variable the study is not valid.

Internal validity refers *the degree to which causality could be inferred*. The findings must accurately describe the phenomena under investigation.

Necessary conditions in for determining causality

- **Temporal antecedence:** the treatment (x), or the cause must come before the effect. X must precede Y.
- **Covariation of the cause and effect:** the cause (X) and effect(Y) must vary together. When our treatment changes, the outcome must also change. Whenever treatment (X) is present the outcome (Y) should also be present.
- **Absence of lack of alternative explanations:** the covariation of x and y does not necessarily show the presence of causality because the relationship between the treatment and outcome could be attributed to other confounding extraneous variables or "the third variable" and "missing variable problem" should be avoided in order to make casual inference. The conditions of relationship and antecedence could be easily ensured through control and other techniques. However, the task of avoiding alternative explanations is very difficult.

A. construct validity

The word construct refers to those variables which could not be directly observed. Construct validity could be considered as a group of different validity types in which we try to answer whether our test, questions, or tools we are using reflects the construct we are trying to measure. The following definitions could be given for construct validity. Construct validity:

- Do our tools, questions equate the concept and definition of our construct?
- Generalization from our tools, program, measure, tests, questions, etc to the concept of the program or test.
- The extent to which our operationalization accurately reflects its construct

Construct validity encompasses different kinds of validity.

a. Convergent validity

Convergent validity refers to the degree to which **the measure is similar to (converge to) other measures it should theoretically be similar**. For example, anxiety measure should converge to constructs for stress.

b. Discriminant validity

Discriminant validity refers to the extent to which the measure or test (questions) *is not similar to other variables it should not be similar*. It shows how **the test inversely varies with a measure of an opposite construct**. For instance, reverse coded items should be negatively correlated to the other questions. As an example, questions for altruistic behaviour should not be correlated to reversely coded aggressive behavioural measures.

- The ability to discriminate between dissimilar constructs.

▪ **Threats to construct validity**

The construct validity could be jeopardized when the tests developed to measure a specific construct could measure more than or less than the intended construct. The following are the main threats listed by Trochim (2006) as follows:

▪ ***Inadequate preoperational explication of constructs***

This problem occurs when we fail to clearly define or articulate the construct before it is transferred into measures or treatments.

▪ ***Mono-operation bias***

Mono-operation bias is directly related to the partial measurement of independent variables. It could also refer to the under representation of the construct when the researcher uses a single version independent variable (*x*, cause) to assess the construct (in order to determine its effect on the dependent variable). For example, the use of only empathetic relationship may not reflect the concept of humanistic psychotherapy.

To solve mono-operation bias, it is determinant to include multiple forms of the independent variable.

▪ ***Mono-method bias***

Mono-operation is related to the independent variable whereas the mono-method bias is related to the partial measurement of the dependent variable. This bias occurs when we measure the part of the dependent variable. For example, measurement of the psychological component of anxiety alone by ignoring the physiological symptoms could lead to the mono-method bias.

The use of multiple measure, pilot or side studies could minimize the mono-method bias.

▪ ***Interaction of treatments***

When our treatments interact with other treatments to produce a change in the dependent variable threat may occur to our construct validity. For example, we studied the effect of discussion and lecture method on memory and conclude that discussion enhances memorization. The score of students in the discussion condition may interact with the students' independent study activities.

- ***Limited generalizability across constructs***

This refers to the "unintended consequences" of the treatments. Unintended consequences could occur when the treatment (x) is effective in causing the behaviour but with the possibility of other negative consequences. For example, a given drug could reduce the anxiety (target variable) but at the same time it produces headache, and abdominal aches.

Therefore, the generalization should be extended from the target variable to other factors.

- ***Level of treatments***

Level of treatment threat occurs when our treatment effect is limited only to the specific level of our independent variable. Slight change in the level of independent variable could be used to curb this problem. When the level of our independent variable is limited to a specific level, our conclusion could be highly affected. For example, to study the effect of chat on academic achievement, we give 20 gm of khat for the experimental group and no khat for the control group; and we conclude that there is no difference between the score of students who use khat and who don't use khat.

To combat, this problem, we can slightly increase and decrease the level of independent variable.

- ***Confounding variables***

Confounding variables are those extraneous variables varying systematically with the treatment (x). For example, a social psychologist studied the effect of physical attractiveness on persuasion and found that more attractive persons are more persuasive as compared to less attractive people. However, the smiling face of the attractive person could have accounted for the persuasion.

B. Face validity

Face validity occurs when test looks (seems) capable to measure what it aims to measure. Face validity is the weakest form of validity. It simply uses **subjective judgment in order to determine validity**. Face validity could or could not be accurate.

C. Content validity

Content validity is ensured when the subject matter covered by the test (construct) contains **representative sample of questions from each content area**. For example, questionnaires developed to measure anxiety should contain representative questions to assess physiological and psychological components of the anxiety symptoms. Here, the determinant factor is the inclusion of relevant contents of our measure to ensure content validity.

D. Criterion Related Validity

Criterion related validity, generally tries to check the test against a predetermined criterion. Simply, criterion validities emphasize on producing results comparable to other similar or same previous studies.

a. Predictive validity

Predictive criterion validity refers to the **use of tests to predict some specific criterion**. The questions used to measure a given variable should be able to predict something they should theoretically predict. For example, the intention to be alone should predict lower level of affiliation. In addition, mathematics score could predict ability of the person for physics, and high school leaving exam results or Aptitude tests should predict college success.

b. Concurrent validity

Concurrent validity is an indicator of a measurement devices' (test, questionnaire) ability to change directly with a measure of the same construct. It is a type of validity where we **compare our test scores with another known standard** (already valid test). Here two measures can be taken at the same time. For example, to examine children's aggression, we can measure the attitude of the parents towards their children's aggression, and the attitude of the teacher towards the children's aggression. Finally, we can correlate the two scores in order to determine the validity. If the correlation is high, we can conclude that the test is valid.

Here is another example; we developed a new tool to measure intelligence. Then we can correlate the pilot result with the Wechsler Adult Intelligence Scale (WAIS).

Threats to Internal Validity

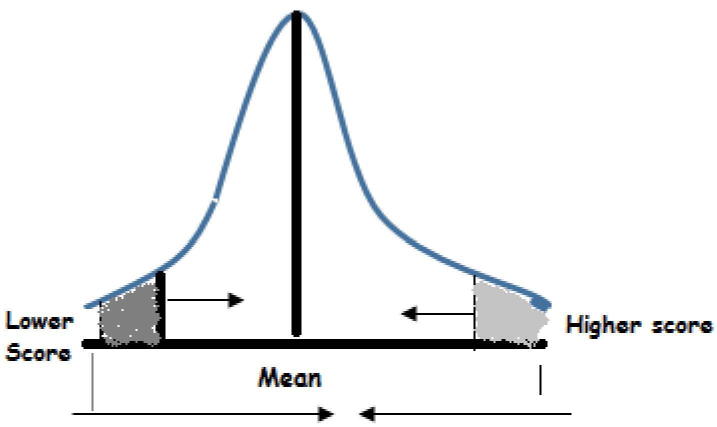
The threats to internal validity refer to all uncontrolled extraneous variables those could be confounded with the effect of the independent variables.

- **History.** History threat occurs when some specific event takes place between testing and measurement of the levels of the independent variables. This refers to historical events those could be confounded with the effect of the treatment (independent variable). In history threat, a specific event occurring between the first and the second measurement of an independent variable such as, big issues as war, big wins or losses, and even minor events like weather and mood change could contribute for history threat. Longer interval of time between the first and second experimental measurement could more probably lead to internal validity threat. History is a threat for one group experiments.

For example, a person wanted to know if the use of cartoon pictures could increase academic achievement. He used no cartoons in the previous year, and the next year he used cartoons. The finding indicated that later students performed better than the former ones. The raised standard (by the MOE) in the next year could have caused the effect. Secular drift, change of long-term social trends could be an example of history threat in longitudinal studies.

- **Maturation.** When the participants grow (older, and get experienced) as the time passes. The participants' growth could be in terms of physical (e.g. hunger, strength, health, etc.) or psychological dimensions (such as, anger, aggression, altruism, etc.) as the time passes. Maturation is confounded with the independent variable when we use only one group pre-post-test design. Maturation becomes detrimental in long term experiments, when participants undergo rapid changes and adjustment. For instance, a researcher could examine the use of humour on pre -schoolers score. The problem of this research could be that the students' ability could increase as their age increases and the humour may not contribute for the effect.
- **Testing.** Multiple testing threat results when the same persons are tested or measured repeatedly in a pre -test or multiple test design. The use of several tests could sensitize the participants to be aware of the research topic and hypotheses. The familiarity with the topic may initiate social desirability (the tendency of participants to be good) and demand characteristics. The score of participants (dependent variable measure) in the post-test could be higher than that of the pre-test in spite of the treatment administered. For example, a researcher measured pre intervention (i.e. public TV campaign) awareness of people about schizophrenia. Then, he measured the same people to know how their awareness changed based on the campaign. The finding indicated that the treatment (TV campaign) has worked. However, the result could be a mere result of the first questionnaire. Testing is a threat for one group designs.
- **Instrumentation.** Instrumentation becomes a threat if the measuring device (observer, scorer, person, machine, or tools) is changed or modified overtime. Researchers may become more accurate in measuring certain behavioural patterns such as, aggression, anxiety, etc. as the experiment continues. The use of exactly similar test or the use of alternative test "equivalent" in type and difficulty could help here.

- **Selection.** When participants in one group (before treatment) are basically different participants in the other group selection threat is more likely to occur. Nonrandom assignment of participants (e.g. based on self-selection criteria) could be confounded with the effect of the treatment. For example, all active participants could be willing to join the experimental group and the rest may compromise the control group. The use of all clever experimental group students could not be attributed to the treatment effect.
- **Mortality (attrition).** Attrition or mortality of subjects refers to differential loss or dropout of participants. Attrition is a threat for designs involving more than one groups. Sometimes, individuals with low scores may drop out and the final result could be attributed to the administered treatment.
- **Statistical regression.** Statistical regression occurs when we select participants on the basis of some extreme scores. For example, we may select students with lower GPAs in order to give them tutorial and based on different reasons, the pretest result could be extremely lower. The next time, in posttest, the score of the students can regress to the normal (mean). Regression to the mean is more likely to occur when we have nonrandom sample from the population and if the two measures are imperfectly related.

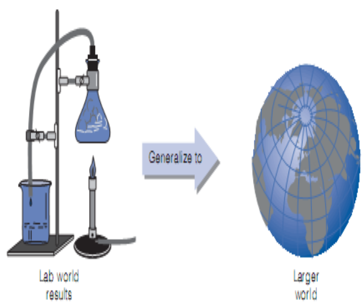


- **Selection interaction:** when the above variables selectively influence specific group. For example, history, maturation, statistical regression, instrumentation, and testing selectively affects participants in (a) specific group.
  - ✓ Selection history threat
  - ✓ Selection testing threat
  - ✓ Selection instrumentation threat
  - ✓ Selection maturation
  - ✓ Selection statistical regression
  - ✓ Selection attrition

2.2.3. External Validity

External validity refers to the degree to which **the study could be generalized to other person in other time.**

External validity shows how the experiment could be applied to the wider population or situations other than the testing environment. For example, how our experiment could relate to the real-life setting.



Generalization models

- ***The sampling models***. To ensure external validity, this model emphasizes identifying the population followed by drawing a representative sample from it. After conducting the experiment with the sample, we generalize the results back to the population.

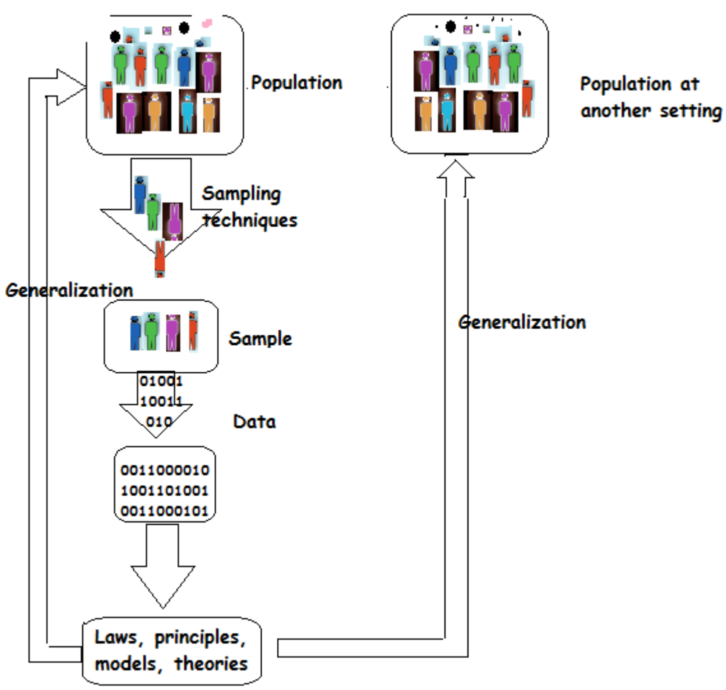


Fig. 2.1.*How external validity works. We study samples and generalize back to the population.*

- ***The proximal similarity model.*** This model assumes generalizing the finding to more similar (proximal) other persons, places, or times of the study. We generalize the result to a more proximally similar time, place, setting, and people of our study.

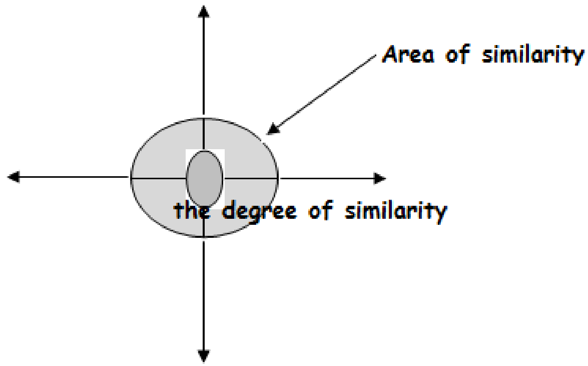
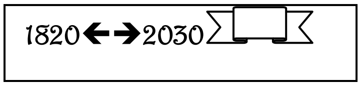


Fig. 2.2. *The role of proximity in generalization*

Types of external validity

- ***Population validity:*** when the result is applied beyond population of the study. Experiments on animals should be generalized to humans for stronger population validity. The results of our experiment should be applied across socio-demographic variables.
- ***Ecological validity:*** refers to a situation when the laboratory experiment could be applied to real life situation. Lab experiments are believed to have high internal validity but low ecological validity. However, field experiments have higher ecological validity compared to the controlled lab experiment.
- ***Temporal validity:*** occurs when the results of the experiment is applicable beyond the time of the research. Theoretically, experiments with higher temporal validity are expected to be generalized to time before and after the experiment.



Threats to External Validity

- ***Reactive testing:*** when pretesting affects participants' response to the independent variable.
- ***Interaction of selection and treatment:*** occurs when the results of the experiment could only be generalized to participants represented in the study. Representation of unique participants (e.g. college students, freshman students, street children, etc.) is typically responsible for interaction of selection effect.

- ***Reactive arrangement***: when the experimental arrangement leads the subjects to think that they are in an experiment.
- ***Multiple treatment arrangement***: occurs when participants are exposed to multiple treatments, exposure to earlier treatment(s) may affect the later treatment(s).

2.2.4. Statistical conclusion validity

Statistical conclusion validity refers to **the degree to which the experimental result is statistically valid.** This is mainly related to the fulfilment of statistical assumptions in the analysis of specific type of experimental design. For example, if the assumption of normality of population distribution (in using ANOVA) is violated the research may lack statistical validity.

2.2.5. Threats in Experimental Research

The social threat to construct validity stems from interpersonal relationships throughout the processes of the research.

A. Demand Characteristics

✓ ***Demand characteristics***. Indicates hypothesis guessing from the experimental situation. Subjects may get hidden clues to guess the objectives, or hypotheses of the research and act or react based on their guess. The participants may guess the hypothesis by involvement in and reading psychology experiments, and information from friends. Martin (2008) stated that the reaction of the students depends based on if they are:

✓ - ***Cooperative participants (good participant effect)***: try to confirm the hypotheses (objectives) of the researcher based on their guess. Cooperative participants try to fulfill the perceived demand (confirming the hypothesis) of the experimenter. The reason of the participants according to Bordens and Abbot (2005) includes desire to help the researcher, desire to help science, and social desirability. Orne (1962) found that participants could cooperate for hours by engaging in repetitive and boring tasks to please the researcher. Martin Orne requested the participants to make 224 additions of random digits on 2000 sheets of paper (followed by an instruction to tear it into more than 32 pieces). They complied until the researcher gave up and told them to stop after five and half hours.

- ***Defensive participants***: act in a way that pleases themselves rather than the experimenter. Defensive participants act based on their own interests. Some participants become defensive because they worry that they are going to be evaluated (measured). Apprehension could severely affect the outcome of the experiment.
- ***Non-cooperative participants***: are negative participants with a tendency to act in a way to contradict the hypotheses of the researcher. Most negative participants try to spoil the experiment, if they are forced to participate.

✓ ***Subject expectancy effect*** (placebo effect). When the expectation of the participants produce the change in the dependent variable because of their belief.

✓ ***Social desirability:*** when participants respond in order to look good in front of the researcher (impression formation) rather than giving genuine responses. Social desirability could also include lying and exaggeration.

✓ ***Hawthorne effect***: The research itself. The research by itself may produce a change in the desired behaviour. The Hawthorne effect is named after the famous study of Elton Mayo in western Electric Plant, Hawthorne in 1930s. In this study, the productivity of workers increased regardless of lightning, rest, pauses, and number of hours worked due to attention of the researchers for the subjects. Similarly, the attention (social relation) of the researchers could bring about a change in the dependent variable.

✓ ***Rosenthal effect***: Experimenter expectation can change the behaviour of the participants. Or the researchers may treat the participants differently based on their expectation.

✓ ***Evaluation apprehension***. When peoples' performance decrease not due to the treatment (x) but because of their hate for being evaluated or tested.

✓ ***Condition diffusion***. Communication between participants from different condition groups during the testing could greatly affect the result of our research.

✓ **Resentful demoralization**. A group that is receiving nothing finds out that a condition (treatment) that others are receiving is effective.

✓ **Diffusion of Treatment**: the independent variable may be diffuse to the control group.

Ways to minimize demand characteristics

- ***Automation***: using technology rather than human beings. For example, the use of tape recorder instruction or computer generated instructions could help to minimize demand characteristics.
- ***Blind and double blind experiments***: when both the experimenter and the participants are unaware of the conditions of the treatments (which participants are in control and which are in experimental group).
- ***Using multiple experimenters***: the use of various experimenters could minimize demand characteristics.
- ***The use of deception***: deliberately misleading the participants from guessing the hypothesis.
- ***Unobtrusive measures:*** measuring participants without their awareness that they are being measured.
- ***Disguised objectives:*** giving hidden reasons very difficult to understand.
- ***Perceived awareness of research hypothesis (PARH):*** asking or scaling the participants if they have been aware about the hypothesis of the study at the end of the experiment. Rubin, et al. (2010) developed the scale containing questions such as: "I knew what the researcher was investigating", "I was unclear about exactly what the researcher was aiming to prove in this research", " I had a good idea about what the hypotheses were in this research", on 7 point scale.

Chapter Three
Experimental Designs

Research designs are indicators of the overall structure of the research. It is a way to carefully plan experiments in advance so that your results are both objective and valid. Experimental design determines the tool, participant selection, number of treatments, number of groups, control, hypotheses, and whether similar or different participants are used in a group, etc.

Ideally your experimental design should:

- Describe how participants are allocated to different conditions.
- Minimize or eliminate confounding factors which can offer alternative explanations for the experimental results.
- Allow you to make inference about the relationship between IDV and DV.
- Reduce variability, to make it easier for you to find difference in treatment outcomes.

Basic experimental design notations:

X=treatment (subjects in the experimental group exposed with it)

O=Observation, testing, and measurement of the outcome

E=Experimental groups

C=Control group

R=Random assignment of participants to different groups

NR=Non-random assignment of participants to different groups

The number of rows indicates the number of control group and "X" refers to subjects in the experimental group.

Mundane and Experimental realism

Experimental realism: the experimental setting that elicits natural response of participants. The participants should not feel and respond in a way that they are in experiment.

Mundane realism: the extent to which the experimental situation corresponds to real life situation.

The degree to which the researcher assigns subjects to conditions and groups distinguishes the type of experimental design.

3.1. True Experimental Design

True experimental design employs:

- ***Randomization***: random selection and assignment of participants to different treatment conditions.

- ***Control***: eliminating or minimizing extraneous variables.
- ***Multiple groups***: the existence of more than one group.
- ***Manipulation***: the application of independent variable over dependent variable.

Types of True Experimental Design

Experimental design mainly involves the use of control group and randomization.

A. Pre-test-Post-test Group Design

- The subjects are randomly assigned to either the experimental or the control group.
- Measurements are taken both before and after a treatment.
- The experimental group receives the treatment while the control group will receive no treatment (in medicine placebo is given).
- Finally, both groups are post-tested to examine the effects of manipulating the independent variable on the dependent variable.

This design could de deprecated as:

E O_1 X 0_2

C 0_3 0_4

For example, the experimental group is tested before and after taking cartoon aided education. The control group is measures twice without taking carton aided education.

In medical research this is termed as *randomized control (clinical) tri*al (RCT). The clinical design could be written as follows in which Y represents placebo.

E O_1 X 0_2

C 0_3 Y 0_4

Analysis: this designed can be analysed by using *independent samples* ***t*-test**.

B. Post-Test Only Design

- The subjects are randomly assigned to either the experimental or the control group.
- Neither group is pretested before the implementation of the treatment.
- The treatment is applied only to the experimental group.
- The post-test is carried out on both groups to assess the effect of the treatment or manipulation.
- This type of design is common when it is not possible to pretest the subjects.

This design can be represented as:

E X 0_1

C 0_2

This experimental design can be calculated by using independent group's *t*-test.

C. Solomon Four Groups Design

This method is really a combination of the previous two methods and is used to eliminate potential sources of error.

- Subjects are randomly assigned into one of four groups.
- There are two experimental groups and two control groups.
- Only two groups (E_1 and C_1) are pretested.
- One pretested group and one un-pretested group receive the treatment.
- All four groups will receive the post-test.

The Solomon four group design can be represented as:

E_1 O1 X O2
C_1 O3 _ O4
E_2 _ X O5
C_2 _ _ O6

This design is very essential in considering external validity. The effect of "X" is replicated and compared in four different conditions. The main drawback of this method is that it requires size of participants, at least 30 in each four condition.

D. Randomized Pre-test-post-test Multiple Treatment Groups Design

This is an experimental design in which multiple treatment groups are used instead of the control group. Even though it can be described in various ways, look at the following notation as an example for two treatment conditions:

E_1 O_1 X_1 O_2
E_2 O_1 X_2 O_2

E. Factorial Designs

Factorial design involves two or more independent variables (called factors) having two or more levels. For example, socio-economic variable can be a factor with three levels (high, medium, and low).

Look at the following 2x3 (two by three) factorial design for gender and physical attractiveness on years of imprisonment for an offense of murder.

		Physical attractiveness		
		Low	Medium	High
Gender	Male	$\bar{X}_{11}=12$	$\bar{X}_{12}=8$	$\bar{X}_{13}=5$
	Female	$\bar{X}_{21}=10$	$\bar{X}_{22}=5$	$\bar{X}_{23}=2$

$\bar{X}_{11}$= Mean imprisonment on very unattractive male defendant

$\bar{X}_{12}$=Mean imprisonment on moderately attractive male defendant

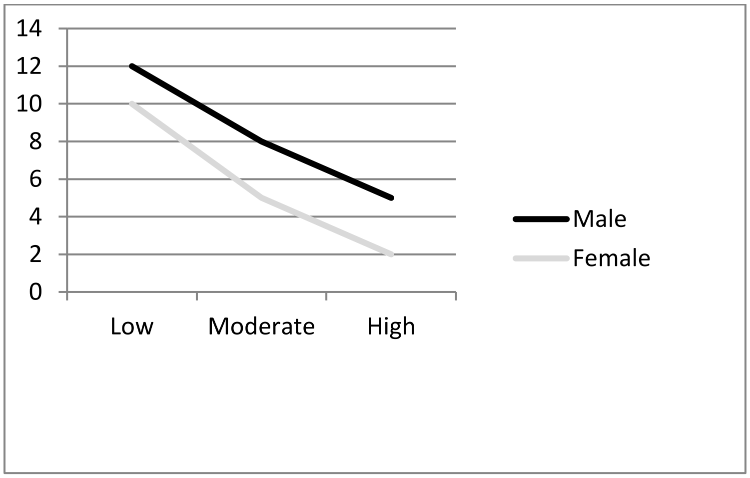
$\bar{X}_{13}$= Mean imprisonment on male highly attractive defendant

$\bar{X}_{21}$= Mean imprisonment on very unattractive female defendant

$\bar{X}_{22}$= Mean imprisonment on moderately attractive female defendant

$\bar{X}_{23}$= Mean imprisonment on highly attractive female defendant

The above hypothetical data could be graphically represented as:



There are two types of effects in factorial design:

Main effect: the independent effect of each factor on the dependent variable.

Interaction effect: the joint effect of factors on the dependent variable. Interaction effect occurs when the effect of one independent variable depend up on the level of another independent variable.

Types of interactions

According to McBurney and White (2007) interaction effects can be categorized as follows:

- *Antagonistic interaction*: when the two variables reverse each other's effects. This shows that there is no main effect of the two variables but there is an interaction. Look at the following graph:
- *Synergistic interaction*: when the two independent variables enhance each other's effect. When the value of one IV is higher, the value of the other also tends to be higher.

3.2. Quasi-Experimental Designs

Quasi experiments are used when true experiments could not be done due to randomization. It has much the same component as a regular experiment, but is missing one or more key components. The three key components of a traditional (true) experimental designs are:

- Pre-post test design
- Manipulation of treatment and control groups
- Random assignment of subjects to groups

Experimenters forced to employ quasi-experimental design due to ethical or methodological reasons. For example:

- Manipulating some treatments (drugs) may be dangerous to human life.
- A regular experiment may be expensive and impossible to fund.
- It might be logistically impossible to control for all variables in a regular experiment

Some experiments naturally fall into groups. An experimenter might want to compare educational experience of first, middle and last born children. Random assignment isn't possible, so this experiments are quasi-experimental by nature.

There are different kinds of quasi experimental designs.

Types of Quasi-Experimental Designs

A. Time Series Design

In time series design, we use series of measurements (tests) before and after the treatment is administered. Time series designs refer to the pretesting and post-testing of one group of subjects at different intervals. The purpose might be to determine long term effect of treatment. Sometimes there is an interruption between tests in order to assess the strength of treatment over an extended time period. When such a design is employed, the posttest is referred to as follow-up.

The design can be shown by the following notation:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

For example:

- The students will be tested several times before and after taking cartoon aided education.
- Traffic fatality rate will be administered before and after the implementation of a crackdown to determine if it reduces the traffic fatality rate in that state.

In the process of analysis average pre-treatment and post treatment scores can be compared by using *t*-test.

B. Equivalent Time Series Design

In the equivalent time series design the intervention is administered several times and the intervention is taken away several times. It is a time series with repeated application of treatment. Observations of the outcome are taken after each administration of the treatment and after each time period where the intervention is taken away.

In notation form we can present this design as:

$$X\ O_1 \quad NX\ O_2 \quad X\ O_3 \quad NX\ O_4 \quad X\ O_5 \quad NX\ O_6$$

Note: NX= no treatment

The stability of the treatment effect can be monitored across treatment administrations, washout periods, and non-treatment periods. Longitudinal effects can also be studied across the multiple observations

C. **Multiple Time Series Design with comparison group**

Multiple time series involves multiple observations or testing before and after treatment (similar to the first type of time series). In addition, the use of multiple measurement of a comparison group that serve as a control group.

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_9 \quad O_{10} \quad O_{11} \quad O_{12}$$

$$O_5 \quad O_6 \quad O_7 \quad O_8 \quad - \quad O_{13} \quad O_{14} \quad O_{15} \quad O_{16}$$

The analysis of this design can also involve the use of averaged observation scores in to analyse independent samples *t*-test in order to determine the effect of the treatment.

D. Pre-test post-test Comparison Group Design

With this design, both a control group and an experimental group is compared, however, the groups are chosen and assigned out of convenience rather than through randomization. We might ask students to participate in a one-semester work experience program. We would then measure all of the students' grades prior to the start of the program and then again after the program. Those students who participated would be our treatment group; those who did not would be our control group.

In notation form we can present this design as:

E	O_1	X	O_2
C	O_1	_	O_2

Note: E= experimental group, C= control group, O=observation

E. The Non-equivalent Group Design

In this exercise you are going to create a non-equivalent group or an untreated control group design of the form

N	O	X	O
N	O		O

where each O indicates an observation or measure on a group of people, the X indicates the implementation of some treatment or program, separate lines are used to depict the two groups in the study, the N indicates that assignment to either the treatment or control group is not controlled by the researcher (the groups may be naturally formed or persons may self-select the group they are in), and the passage of time is indicated by moving from left to right. We will assume that we are comparing a program and comparison group (instead of two programs or different levels of the same program).

This design has several important characteristics. First, the design has pre-test and post-test measures for all participants. Second, the design calls for two groups, one which gets some program or treatment and one which does not (termed the "program" and "comparison" groups respectively). Third, the two groups are non-equivalent, that is, we expect that they may differ prior to the study. Often, non-equivalent groups are simply two intact groups which are convenient to the researcher (e.g., two classrooms, two states, two cities, two mental health centers, etc.).

3.3. Within and Between Subjects Design

3.3.1 Within subjects Design (*Repeated Measures Design*)

A within subjects design is an experiment in which the same group of subjects serves in more than one treatment. Only one group of subjects; subjects receive all levels of the independent variable at different times. In this condition all/each participant will be exposed to all kinds of treatment. For example, the same person can be exposed to cartoon aided education, discussion method, and lecture methods. In addition the participants can be measured after each treatment.

This design is helpful in eliminating confounding variables associated to subject variables by using the same participants across each level of the independent variable.

Moreover, within subject design requires smaller sample size (because each person receives treatment in each condition). Between subjects design is inconvenient for experiments involving specific people with particular type of personality, disability, disorder, etc. In such conditions, within subjects designs are very important.

Types of Within Subjects Design

Repeated measures design can be categorized as follows:

A. Single factor within-subjects Designs

Single factor within-subjects design involves experimentation with only one independent variable having different levels. Participants assigned to different levels of an independent variable on random base. Each participant is exposed to only one level of an independent variable. This design has two variants:

- ***Single factor two levels design***

Single factor two levels design exposes the same person (participant) to two levels of a single independent variable (e.g. 25 gm, 50 gm of a drug). The response of the participants is measured after and before the treatment. For example, a psychologist can measure the effectiveness of chewing chat at different level on academic score.

Look at the following to understand this:

Condition	Treatment1 (25 gm chat)	treatment 2 (50 gm chat)
S1	X_1	X_2
S2	X_1	X_2
S3	X_1	X_2

- ***Single factor multi-level design***

Single factor multi-level design is used when a single IV consists more than two levels. The levels can three, four, five, etc. For example second year psychology students can be exposed to different levels of a given drug (e.g. 25 gm, 50 gm, and 100 gm).

For example, look at the following notation with example:

R X_1 O_1 = 25 gm

R X_2 O_2= 50 gm

R X_3 O_3 =100 gm

Subject	Treatment 1(25 gm)	Treatment 2 (50 gm)	Treatment 3 (100 gm)
S1	$\bar{X}_1$	$\bar{X}_1$	$\bar{X}_1$
S2	$\bar{X}_2$	$\bar{X}_2$	$\bar{X}_2$
S3	$\bar{X}_3$	$\bar{X}_3$	$\bar{X}_3$

Where: $\bar{X}_1$=Mean of treatment 1 $\bar{X}_2$=Mean of treatment 2 $\bar{X}_3$=Mean of treatment 3

B. Multiple factor within-Subjects Design: Within-Subjects Factorial Design

This is used when there are two and more independent variables with different levels (e.g. time spent on studying and motivation). Here, subjects are exposed to every combination of every treatment. In within-subjects factorial design, the subjects are assigned to each conditions (each level of the treatment) of the experiment on a random basis. For example, if we are looking to investigate the effect of stress and motivation on students' academic score, the first factor is stress and the second factor is motivation.

For example, look at the following 2x3 factorial within subjects design towards the effect of stress and motivation on academic score.

Factor A: stress

- Level: A1: Low
  A2: High

Factor B: Motivation

- Level: B1: Low
  B2: Moderate
  B3: High

			Stress	
		Subjects	A1 (low)	A2 (High)
Motivation	B1 (low)	S1 S2 S3 S4 S5	Condition 1	Condition 2
	B2 (moderate)	S1 S2 S3 S4 S5	Condition 3	Condition 4
	B3 (high)	S1 S2 S3 S4 S5	Condition 5	Condition 6

In this design, a participant is exposed to each condition on a random basis. This means, the participant will be measured six times in each treatment in high, moderate and low motivation level accompanied by low and high level of stress.

Advantages and disadvantages of within-subjects design

Advantages

- Requires relatively smaller sample size as compared to between-subjects design. It does not require a large pool of participants.

- Reduce errors associated with individual difference. All individuals are exposed to all levels of a treatment, so individual difference will not distort the result.

Disadvantages: *major threats in within-subjects design*

- ***Carryover effect***: when the prior treatment(s) cause a change in the DV scores of the participants due to the subsequent treatment. This is directly related to the practice effect manifested in as:
  - ***Learning***: DV score increases due to practice.
  - ***Fatigue***: participants may become exhausted, bored, or simply uninterested after taking part in multiple treatments or testes.
  - ***Sensitization***: the first treatment may alert the participants in the second condition.
  - ***Contrast***: extremely contrasting situations can elicit different responses. For example, watching attractive photograph followed by unattractive photo may lead the participants to guess the hypothesis of the experiment.
  - ***Adaptation***: adjustment to the treatment condition may occur as the treatment continues.
- ***Attrition***: subjects may leave the experiment after participating in the first condition.

  Look at the following example of attrition:

Order of X	**Subject**	**X1**	**X2**	**X3**
1	S_3	4	3	4
2	S_1	2	--	2
3	S_5	3	4	--
4	S_2	5	--	3
5	S_4	3	5	5

-- =missing DV score due to attrition
X =treatment
S= Subject

- ***Order effect***: when the DV score varies based on the presentation order of the treatments.
- ***Sequence effect***: the arrangement of the treatments can be a source of variability in repeated measures design.

3.3.2 Between Subjects Design

In between-subjects design, different groups of participants are assigned to different treatment conditions based on randomization. Some participants are exposed to one level of the IV and others to another level. Different participants are exposed to different conditions of the experiment.

Steps of conducting between subjects design

Bordens and Abott (2005) proposed the following basic steps in conducting between subjects design:

- Selecting sample from the population
- Random assignment: assignment of participants to different conditions based on chance. One participant is exposed to only one condition.
- Giving treatments to different group of participants
- Measurement of the DV score
- Comparison: comparing the mean DV scores of different participants in different group.

When we conduct experiment by using between subjects design, we try to control extraneous variables by using matching.

- **Precision matching**: matching based on identical scores. People will be matched based on precise score on certain attributes (e.g. anxiety, depression, aggression, academic score, etc.).
- **Range matching**: is used when we are unable to use precise matching. Here we match participants based on range scores (e.g. we can match students based on the following GPA score; 1.5-1.9., 2.0-2.5, 2.6-2.9, etc.).
- **Rank ordered matching**: is used when participants are ordered based on their scores (e.g. 1.2, 1.2, 1.3, 2.0, 2.3, 2.4…3.9, 4.0).

3.4. Controlling in Experimental research

Extraneous variables are controlled through:

A. Elimination

Elimination involves taking out or *removing all extraneous variables*. For example, if we are studying the effect of using cartoon pictures on students GPA. The class conditions such as room size, colour of the room, noise outside the classroom, etc. can affect the outcome of the experiment. Here the experimenter can eliminate the noise of other students by posting notices not to disturb or appointing others to protect others from coming to the experimental area. Elimination is more theoretical and it is impractical to eliminate all extraneous variables.

B. Constancy

Constancy is used where elimination is impossible. Constancy involves *keeping all treatment conditions approximately similar*. For example, keeping room temperature, room size, number of students, weather, and room colour approximately similar indicates the method of constancy.

C. Balancing

Balancing is used when we fail to use elimination and constancy. Balancing refers to the act of *distributing the effect of the extraneous variable across conditions*.

Practically we can identify our extraneous variable (e.g. gender), randomize (assign participants based on gender; half male and half female), and randomly assign participants from the groups to different treatment conditions.

D. Double blind experiment

The other way to overcome the influence of extraneous variables is by using double blind experiment in which both the participants and the experimenter do not know which participants are assigned to specific group.

E. Matching: Control for individual difference by matching similar subjects or groups with each other.

Chapter Four
Reporting Experiments
Elements of Experimental research
Typical format of a journal article

Almost all journal articles dealing with research studies are divided into different sections by a means of headings and sub heading
gs. Although there is variation among journals with respect to the terms used for the headings and subheadings, there is relatively standard format for published articles.

Typical journal article can possibly contain the following format:

1. Abstract
2. Introduction
   a. Background
   b. Statement of purpose
   c. Hypothesis
3. Method
   a. Participant
   b. Measure
   c. Procedure
   d. Statistical plans
4. Result
5. Discussion
6. References

1. Abstract

Summarizes the entire research study and appears at the beginning of the article. Although it normally contains fewer than 150 words, the abstract usually provides the following information:

I. A statement of the purpose or objective of the investigation,
II. A description of the individuals who served as participants,
III. A brief explanation of what the participants did during the study, and
IV. A summary of the important findings.

The sole purpose of the abstract is to provide readers with a brief overview of the study's purpose, methods, and findings. Thus most abstract indicates why the study was conducted, how the researcher went about trying to answer the questions of interest, and what was discovered after the study's data were analysed.

2. **Introduction**: The introduction of an article usually contains two items: a description of the study's background and a statement of purpose. Sometimes a third portion of the introduction contains a presentation of the researcher's hypotheses.

a. **Background**

Most authors begin their articles by explaining what caused them to conduct their empirical investigations. Perhaps the author developed a researchable idea from discussions with colleagues or students. Maybe a previous study yielded unexpected results, thus prompting the current researcher to conduct a new study to see if those earlier results could be replicated. Or, maybe the author wanted to see which of two competing theories would be supported more by having the collected data conform to its hypotheses. By reading the introductory paragraph(s) of the article, you learn why the author conducted the study.

In describing the background of their studies, authors typically highlight the connection between their studies and others' previously published work. Whether this review of literature is short or long, its purpose is to show that the current author's work has been informed by, or can be thought of as an extension of, previous knowledge. Such discussions are a hallmark of scholarly work. Occasionally, a researcher conducts a study based on an idea that is not connected to anything anyone has investigated or written about; such studies, however, are rare.

Note that the background section focus on what has been studied and found in earlier research investigations instead of presenting opinion, hope, or subjective experiences. Moreover, this part of the researcher's report is characteristic of published articles, doctoral dissertations, master's theses, and reports from independent and government research agencies.

b. **Statement of purpose**

After discussing the study's background, an author usually states the specific purpose or goal of the investigation. This statement of purpose is one of the most important parts of a research report, because in a sense, it explains what the author's "destination" is. It would be impossible for us to evaluate whether the trip was successful—in terms of research findings and conclusions—unless we know where the author was headed. The statement of purpose can be as short as a single sentence or as long as one or two full paragraphs. It is often positioned just before the first main heading of the article, but it can appear anywhere in the introduction. Regardless of its length or where it is located, you will have no trouble finding the statement of purpose if a sentence contains the words, "the purpose of this study was to . . ." or "this investigation was conducted in order to . . ."

c. Hypothesis

After articulating the study's intended purpose, some authors disclose the hypotheses they had at the beginning of the investigation. Other authors do not do this, either because they did not have any firm expectations or because they consider it unscientific for the researcher to hold hunches that might bias the collection or interpretation of the data. Although there are cases in which a researcher can conduct a good study without having any hypotheses as to how things will turn out, and although it is important for researchers to be unbiased, there is a clear benefit in knowing what the researcher's hypotheses were. Simply stated, outcomes compared against hypotheses usually are more informative than are results that stand in a vacuum.

3. Method

In the method section of a journal article, an author explains in detail how the study was conducted. Ideally, such an explanation should contain enough information to enable a reader to replicate (i.e., duplicate) the study. To accomplish this goal, the author addresses three questions:

- Who participated in the study?
- What kinds of measuring instruments were used to collect the data? and
- What were the participants required to do?

The answer to each of these questions is generally found under an appropriately titled subheading in the method section.

a. Participants

Each of the individuals (or animals) who supplies data in a research study is considered to be a participant or a subject. (In some journals, the abbreviations S and Ss are used, respectively, to designate one subject or a group of subjects.) Within this section of a research report, an author usually indicates how many participants or subjects were used, who they were, and how they were selected. A full description of the participants is needed because the results of a study often vary according to the nature of the participants used. This means that the conclusions of a study, in most cases, are valid only for individuals (or animals) who are similar to the ones used by the researcher. For example, if two different types indicate whether the participants were high school students, adults, patients in a mental hospital, or whatever. What works for a counselor in a mental hospital may not work at all for a counselor in a high school (and vice versa).

It is also important for the author to indicate how the participants were obtained. Were they volunteers? Were they randomly selected from a larger pool of potential participants? Were any particular standards of selection used? Did the researcher simply use all members of a certain high school or college class? of counseling techniques are compared and found to differ in terms of how effective they are in helping clients clarify their goals, it is imperative that the investigator.

b. Measures

This section of a journal article is normally labeled in one of five ways: measures, equipment, apparatus, instruments, or materials. This part of the article contains a description of the things (other than the participants) used in the study. The goal here is to describe what was done with sufficient clarity so others could replicate the investigation to see if the results remain the same.

Suppose, for example, that a researcher conducts a study to see if males differ from females in the way they evaluate various styles of clothing. To make it possible for others to replicate this study, the researcher must indicate whether the participants saw actual articles of clothing or pictures of clothing (and if pictures, whether they were prints or slides, what size they were, and whether they were in color), whether the clothing articles were being worn when observed by participants (and if so, who modeled the clothes), what specific clothing styles were involved, how many articles of clothing were evaluated, who manufactured the clothes, and all other relevant details. If the researcher does not provide this information, it is impossible for anyone to replicate the study.

Often, the only material involved is the measuring device used to collect data. Such measuring devices—whether of a mechanical variety (e.g., a stopwatch), an online variety, or a paper-and-pencil variety (e.g., a questionnaire)—ought to be described very carefully. If the measuring device is a new instrument designed specifically for the study being summarized, the researcher typically reports evidence concerning the instrument's technical psychometric properties. Generally, the author accomplishes this task by discussing the reliability and validity of the scores generated by using the new instrument.1 Even if an existing and reputable measuring instrument has been used, the researcher ought to tell us specifically what instrument was used (by indicating form, model number, publication date, etc.).

One must know such information, of course, before a full replication of the study could be attempted. In addition, the researcher ought to pass along reliability and validity evidence cited by those who developed the instrument. Ideally, the authors ought to provide their own evidence as to the reliability and validity of scores used in their study, even if an existing instrument is used.

c. Procedure

How the study was conducted is explained in the procedure section of the journal article. Here, the researcher explains what the participants did—or what was done to them—during the investigation. Sometimes an author even includes a verbatim account of instructions given to the participants.

To permit a reader to replicate a study, the author must outline clearly the procedures that were followed, providing answers to questions such as (1) Where was the study conducted? (2) Who conducted the study? (3) In what sequence did events take place? and (4) Did any of the subjects drop out prior to the study's completion?

d. Statistical plans

Most research reports contain a paragraph (or more) devoted to the plans for statistically analysing the study's data. In some reports, this information is presented near the end of the method section; in other reports, a discussion of the statistical planof- attack is positioned at the beginning of the report's results section.

4. Results

There are three ways in which the results of an empirical investigation are reported. First, the results can be presented within the text of the article—that is, with only words. Second, they can be summarized in one or more tables. Third, the findings can be displayed by means of a graph (technically called a figure). Not infrequently, a combination of these mechanisms for reporting results is used to help readers gain a more complete understanding of how the study turned out.

5. Discussion

The results section of a journal article contains a technical report of how the statistical analyses turned out, whereas the discussion section is usually devoted to a nontechnical interpretation of the results.

In other words, the author normally uses the discussion section to explain what the results mean in regard to the central purpose of the study. The statement of purpose, which appears near the beginning of the article, usually contains an underlying or obvious research question; the discussion section ought to provide a direct answer to that question.

In addition to telling us what the results mean, many authors use this section of the article to explain why they think the results turned out the way they did. Although such a discussion occasionally is found in articles where the data support the researchers' hunches, authors are much more inclined to point out possible reasons for the obtained results when those results are inconsistent with their expectations. If one or more of the scores turn out to be highly different from the rest, the researcher may talk about such serendipitous findings in the discussion section.

Sometimes an author uses the discussion section to suggest ideas for further research studies. Even if the results do not turn out the way the researcher anticipated, the study may be quite worthwhile in that it might stimulate the researcher (and others) to identify new types of studies that need to be conducted. Although this form of discussion more typically is associated with unpublished master's theses and doctoral dissertations, it occasionally is encountered in published forms of research reports.

It should be noted that some authors use the term conclusion rather than discussion to label this part of the research report. These two terms are used interchangeably. It is unusual, therefore, to find an article that contains both a discussion section and a conclusion section.

6. References

A research report normally concludes with a list of the books, journal articles, and other source material referred to by the author. Most of these items were probably mentioned by the author in the review of the literature positioned near the beginning of the article.