

# **RESEARCH METHODS & EXPERIMENTAL DESIGN**

**A set of notes suitable for  
seminar use**

by Robin Beaumont

Last updated: Sunday, 26 July 2009

e-mail: [robin@organplayers.co.uk](mailto:robin@organplayers.co.uk)

**This handout is part of a course. For details please see Section 7.1 at:  
<http://www.robin-beaumont.co.uk/virtualclassroom/contents.htm>**

# Contents

<b>1.</b>	<b><i>Introduction</i></b>	<b>3</b>
<b>2.</b>	<b><i>Common Components of Research</i></b>	<b>4</b>
<b>3.</b>	<b><i>Approaches to Research</i></b>	<b>5</b>
3.1	<b>Classification one - Qualitative / Quantitative</b>	<b>5</b>
3.2	<b>Classification two - Deductive / Inductive</b>	<b>6</b>
3.3	<b>Classification three - Mixed</b>	<b>7</b>
<b>4.</b>	<b><i>Experiments, Quasi Experiments and Non-Experiments</i></b>	<b>8</b>
4.1	<b>Experiments</b>	<b>8</b>
4.1.1	Pretest Post-Test Control Group Design	9
4.1.2	Post-Test only Control Group Design	9
4.1.3	Solomon Four-Group Design	9
4.1.4	Factorial Design	10
4.1.5	4.1.5 Some factors to Consider	10
4.2	<b>Quasi-Experiments [Field Methods]</b>	<b>11</b>
4.2.1	Interrupted Time-Series Design	11
4.2.2	Multiple Time-Series Designs	11
4.2.3	Non-equivalent Control Group Design	12
4.2.4	Separate-Sample Pretest-Posttest Design	12
4.2.5	Separate Sample Pretest-Posttest Control Group Design	12
4.3	<b>Non-experimental Designs</b>	<b>13</b>
4.3.1	One-Shot [Cross-sectional] Study [Case Study][Survey]	13
4.3.2	One-Shot Correlational Study	13
4.3.3	One-Group Pretest-Posttest Design	14
4.3.4	Static-Group Comparison Design	14
<b>5.</b>	<b><i>Reliability</i></b>	<b>15</b>
5.1	<b>Testing Reliability</b>	<b>15</b>
5.2	<b>Validity (part 1)</b>	<b>15</b>
<b>6.</b>	<b><i>Operationalisation</i></b>	<b>16</b>
6.1	<b>Concepts</b>	<b>16</b>
6.1.1	Dimensions of a Concept	16
6.2	<b>Measures</b>	<b>17</b>
6.2.1	Variables & Constants	17
6.2.2	Scales	18
6.3	<b>Graphical Method of Representing Operationalisation</b>	<b>19</b>
<b>7.</b>	<b><i>Validity (part 2)</i></b>	<b>20</b>
7.1.1	Relationship with Reliability	20
7.1.2	Types of Validity	20
7.1.3	Face (logical), Content and Construct Validity of Measurements	20
7.2	<b>Importance of Good Operationalisation</b>	<b>21</b>
7.3	<b>Context Specific</b>	<b>21</b>
7.4	<b>Summary</b>	<b>21</b>
<b>8.</b>	<b><i>Hypotheses</i></b>	<b>22</b>
8.1	<b>The Research Hypothesis</b>	<b>22</b>
8.2	<b>The Statistical Hypotheses</b>	<b>23</b>
8.2.1	Null Hypotheses( $h_0$ )	24
8.2.1.1	Testable Hypotheses	24

8.2.1.2	Why 'Null'	25
8.2.2	Relationship between a Inferential Statistic and a Null Hypothesis	26
8.2.3	Alternative Hypothesis ( $h_1$ )	27
<b>9.</b>	<b><i>Validity (part 3)</i></b>	<b>28</b>
9.1	<b>Internal Validity (of Findings)</b>	<b>28</b>
9.2	<b>Validity of Statistics</b>	<b>30</b>
9.3	<b>Construct Validity of Measurements</b>	<b>31</b>
9.4	<b>External Validity [Generalisations]</b>	<b>32</b>
9.5	<b>Other types of Validity</b>	<b>32</b>
9.6	<b>Resume</b>	<b>33</b>
<b>10.</b>	<b><i>Experimental Control</i></b>	<b>35</b>
10.1	<b>Manipulation of Variables</b>	<b>35</b>
10.2	<b>Random Assignment [Allocation]</b>	<b>36</b>
10.3	<b>Control / Placebo Groups</b>	<b>37</b>
10.4	<b>Protocols</b>	<b>37</b>
10.5	<b>Blinding</b>	<b>38</b>
10.6	<b>Controlling / Measuring Inter-subject differences:</b>	<b>38</b>
10.6.1	Homogenous Subjects	38
10.6.2	Blocking	38
10.6.3	Matching	38
10.6.3.1	Problems with Repeated Measures Designs	38
10.6.3.2	Controlling Problems with Repeated Measures Designs	39
10.6.4	Analysis of Covariance (ANCOVA)	39
<b>11.</b>	<b><i>Evaluating Research Reports</i></b>	<b>40</b>
11.1	<b>Robin Beaumont 1995</b>	<b>40</b>
11.2	<b>Phillips L R 1986</b>	<b>41</b>
11.3	<b>Oxman A D 1994</b>	<b>41</b>
11.4	<b>Hawthorn P J, 1983</b>	<b>42</b>
11.5	<b>Greenhalgh 1997</b>	<b>44</b>
<b>12.</b>	<b><i>Questionnaire Design an Introduction</i></b>	<b>45</b>
<b>13.</b>	<b><i>References</i></b>	<b>45</b>
<b>14.</b>	<b><i>Index</i></b>	<b>47</b>

# 1. Introduction

The main aims of this handout are:

- To provide a source of background material which will help you to take part in, and follow the seminar discussions.
- Act as a reference for you during the rest of your course.

Almost all of the material presented in this handout is DEBATABLE. Your views and criticisms, both whilst you are reading this handout, and during the seminars should be written in alongside the relevant material. Just because it has a reference does not mean it is holy!

Many topics, such as validity, reappear several times each time discussing the concept in more depth, this is deliberate as I have found that teaching many of these topics requires us to re-visit and reflect many times before a detailed understanding is gained. I hope this approach also works for you.

The symbols [] are used to provide equivalent or similar names e.g. man[male]

## 2. Common Components of Research

Nearly all research situations consist of the following:

- 1      Experimenter / Investigator / Researcher
- 2      Subject (= Client / Volunteer / Cohort / Group) / Sample
- 3      Data (= Primary or Secondary data)
- 4      Instrument (e.g. thermometer, ergometer, ECG, questionnaire)

In addition the following may also be present:

- 5      Treatment (= Intervention / Manipulated variable / Test group)
- 6      Controls (= Non test group / placebo )

### Exercises

- 1      Would a one-off survey possess components 5 and / or 6? What are the advantages and disadvantages of this?
- 2      Which components do you think a classic laboratory experiment possesses?
- 3      A single blind experimental design is one where the subjects do not know if they are receiving the treatment or a placebo. What is a double blind experimental design?
- 4      What might be the advantages and disadvantages of 'blinding'?
- 5      Devise a diagram to try to demonstrate the possible relationships between the above components.
- 6      What might be the advantages and disadvantages of supplying a placebo treatment to a control group?

During this course we will be looking closely at all the above aspects except data which is covered in the statistical notes. However if you want an introduction about what you should consider when collecting data see: <http://www.robin-beaumont.co.uk/virtualclassroom/chap5/s2/data1.pdf>

## 3. Approaches to Research

There are any different ways of classifying research designs, none of which are entirely satisfactory.

### 3.1 Classification one - Qualitative / Quantitative

#### **Qualitative Methodology Characteristics (a biased view of Qualitative methodologies?)**

- Free to devise own method within basic philosophical approach
- Unstructured - often deliberately no hypothesis e.g. Grounded theory ('investigation / Case Study')
- Interpretation / Subjectivity considered important (reliability / validity ↓)
- Naturalistic (outside the 'laboratory' conditions of quantitative research)
- Aims to provide supporting information ('issues raised') which is individually focused rather than considering the individual as a representative of a sample.
- Causation, association and other possibly static universals are considered to be of less importance than the individuals' experience.

E.g. someone spending time with a boxing club and writing up 'what it was like' (Ethnographic) after getting to know the several of the staff and members<sup>1</sup>

#### **Quantitative Methodology**

- Rigorous – clearly defined rules based upon previous results
- Statistical tests / and data analysis have 'rules'
- Considers Objectivity as being of great importance
- Aims to reject a (null) Hypothesis in its most basic form

E.g. someone spending time in a boxing club, with a specific aim to investigate a specific topic (e.g. the level of health in participants) which probably will have a number of academic papers already available. The study might include using well developed questionnaires specific for this area, which would allow statistical comparison with other groups.

Some people disagree with the above divisions because, they say:

1. Both methods are structured ('Structured analysis') its just different<sup>2</sup>
2. Often researchers use more qualitative techniques in initial research designs and then go on to use more quantitative methods latter on once they have discovered what they want to investigate.

Note: The above has nothing to do with the difference between Qualitative (=Nominal[binary][count][dichotomous]) data and Quantitative (= ordinal / Interval / ratio) data.

#### **Exercises**

1) Consider which approach you might use for the following research questions.

- a. Investigate the feelings people have towards the local sports club, or self help groups for weight reduction.
- b. Investigate the difference between several training regimes in terms of improving anaerobic performance. Or investigate the difference between several training regimes for increasing fitness after a Myocardial Infarction.
- c. Gain some insight into bodybuilders' or doctors perceptions of themselves
- d. Discover if participation in games at school subsequently reduces the chance of ant- social behaviour or incidence of obesity?

2) Can you think of situations where an initial qualitative method may be abandoned for a quantitative one eventually?

For more details about the qualitative / quantitative continuum along with the philosophical assumptions each makes see: [http://www.robin-beaumont.co.uk/virtualclassroom/chap5/s5/comm\\_theories/qual QUAN1.pdf](http://www.robin-beaumont.co.uk/virtualclassroom/chap5/s5/comm_theories/qual QUAN1.pdf)

<sup>1</sup>For a good example of such an approach see John Sugden 'The exploitation of disadvantage: the occupational sub-culture of the boxer' in sport, leisure and social relations John Horne, David Jary, Alan eds. Tomlinson pub. 1993 Routledge & Kegan Paul p.187 - 209

<sup>2</sup>See Nixon H.L 1991? Sports sociology that matters: Imperatives and Challenges for the 1990s Sociology of sport jour.8 281-294 and Harris J C 1989? Suited up and stripped down: perspectives for sociocultural sport studies Sociology of sport jour. 6 335-347



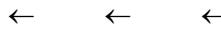
### 3.3 Classification three - Mixed

The way most people work<sup>3</sup>

Select Problem area



Literature Review + find out what's going on  
 Select topic area



Formulate a Hypothesis

What do you intend to test? What is the relationship between the variables (difference, associated / causation etc.) How many are there Develop null and alternative Hypotheses



Design Research



Carry out Research



Interpret Results



Draw Conclusions

Which hypothesis can I reject (refute) How important is the finding? How does it relate to other research?



Report Findings->

- . ->



The above approach usually attempts to reject a particular statement, the null hypothesis, rather than prove it is true. This approach was formalised by Karl Popper (b. 1902) and called it the hypothetico-deductive method. In essence the aim is to gradually reduce the number of possible rival explanations. In contrast to this approach of 'falsifying' hypotheses the alternative traditional approach was to 'verify' a proposition by gathering data (inductivism). Poppers arguments for his approach can be found in 'The logic of scientific discovery' (English translation 1959)

It is fair to say that all the above classifications are too simplistic to represent the real situation because:

- Differing beliefs researchers possess
- Differing ways people work
- Differing environments
- Differing Instruments
- ??????

#### Exercises

- 1 Discuss what are the problems with the above approach?
- 2 Consider a topic area (e.g. one of the following: perceived exertion, psychosis, gastric pain) and investigate from a historical perspective how it relates to the above approach.

<sup>3</sup>Loosely based on a figure in: Sociology by A Giddens 1989 p.663 Cambridge Policy Press in association with Basil Blackwell.

## 4. Experiments, Quasi Experiments and Non-Experiments

The above terms refer to the way a particular piece of research is undertaken. Campbell and Stanley 1966 classified research designs into three types:

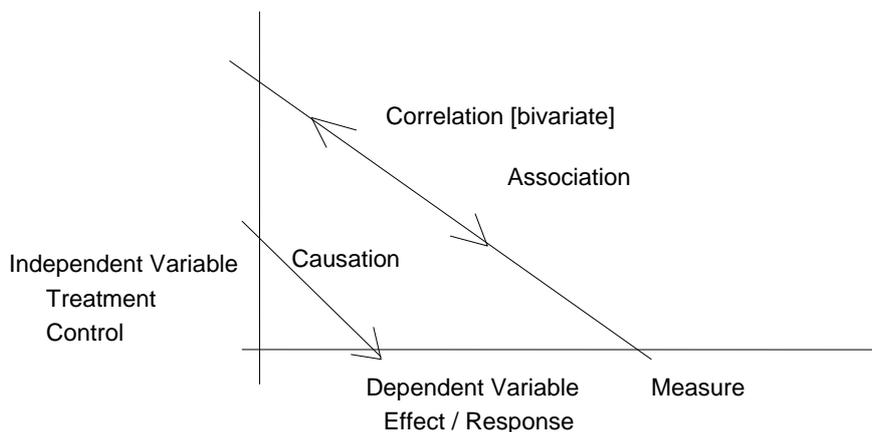
- Experiments
- Quasi Experiments
- Non-Experiments

There are numerous other classifications, but the advantage of this one is that it helps assess the validity of the findings (defined latter). We will consider each type of design in turn.

### 4.1 Experiments

Experimental designs offer the best method available to researchers to be able to investigate causality due to the high degree of control. However such strict control has associated with it its own problems.

Subjects are always both randomly allocated (**Randomisation**) into the different groups and randomly selected (sampled).



Because of the degree of control necessary 'Naturalism' is sometimes reduced. And some people argue that Naturalism is a necessary condition for generalisation [External Validity]. This is explained latter.

Note: Different people use Correlation to mean different things. e.g. "A correlation problem considers the joint variation of two measurements, neither of which is restricted by the experimenter...A regression problem considers the effect upon one variable when another is held fixed at each of several levels"<sup>4</sup> For a quotation from the person who developed the concept of correlation - Francis Galton - see <http://www.stat.ucla.edu/textbook/bivariates/correlation/quotations.html>. and <http://www.amstat.org/publications/jse/v14n1/symanzik.html>

#### Exercise

What do you understand by the term Correlation?

**Key Point:**  
Use the terms "independent variable" and "dependent variable" only with experimental research. With nonexperimental research use "predictor variable" and "criterion variable."

<sup>4</sup>Dixon W J Massey F J. 1983 Introduction to Statistical Analysis p.209 4th ed. international student edition. McGraw-Hill Book Co. London.

### 4.1.1 Pretest Post-Test Control Group Design

This is the classic experimental design

Time 1	Time 2	Time 3
Pretest	Experimental Treatment	Posttest
Pretest	No Treatment (Controls)	Posttest

The controls may be given a placebo. The principle problem concerned with this design is 'subject reactivity to pre-testing' e.g. insertion of a core thermometer may increase stress and cause possibly an unwanted reaction in the subject e.g. possibly raise core temperature? People often consider this problem the same as the "Hawthorne effect", however it does have subtle differences. See the following section on the Hawthorne effect for more details.

**Question:**

What is the advantage of giving a placebo to a group rather than nothing? Hint: Think about alternative therapies.

### 4.1.2 Post-Test only Control Group Design

This consists of the following design:

Time 1	Time 2	Time 3
-	Experimental Treatment	Posttest
-	No Treatment (Controls)	Posttest

The above design removes the problem of pre-testing reactivity but the question then is, can you assume your groups are representative?

### 4.1.3 Solomon Four-Group Design

Group	Time 1	Time 2	Time 3
1	Pretest	Experimental Treatment	Posttest
2	Pretest	No Treatment (Controls)	Posttest
3		Experimental Treatment	Posttest
4	-	No Treatment (Controls)	Posttest

The design allows the measurement of the effects of pre-testing by investigating the difference between post test results of groups 2 and 4.

### 4.1.4 Factorial Design

All the above experimental designs only test the effect of one intervention. It is easy to design (but not so easy to carry out?) extensions to the above catering for two factors. e.g.

#### One factor design

e.g. investigating the effect weight training has on the size of the Gastrocnemius muscle.

#### Two factor design

e.g. Investigate the effect weight training and a special diet has on the size of the Gastrocnemius muscle.

The table below shows how you would model this considering a 'classical' experimental design.

Group	Time 1	Time 2	Time 3
1	Pretest	Diet + training	Posttest
2	Pretest	Diet	Posttest
3	Pretest	training	Posttest
4	Pretest	No Treatment (Controls)	Posttest

### 4.1.5 Some factors to Consider

When considering an experimental design it is worthwhile considering the following:

<b>Operationalisation</b>	What are you going to measure Relationship between the measure(s) and the concept
<b>Instrumentation</b>	How are you going to measure it Questionnaires, Observation, Transducers, Equipment? Reliability?
<b>Sample</b>	Size, Type, How are you going to obtain it, Suitability Typicality. Possible problems with drop outs or compliance.
<b>Groups [units]</b>	One, two or multiple groups Homogeneity Stratification? Control(s)? Replication?
<b>Data collection schedule</b>	Pre Testing? When? How often Serial Measures? When? How often Post Testing? When? How often, Blinding?

#### Exercises:

1. Some of the words in the above sentences will be unfamiliar to you look them up and make notes.
2. What additional factors do you think it would be worthwhile considering from a more practical point of view?

## 4.2 Quasi-Experiments [Field Methods]

These are often called observational studies.

Problems with true experiments:

- a. Artificiality (lack of naturalism)
- b. Reactivity (Subject and Researcher)
- c. Deception?

Unlike the above experimental designs in quasi experiments the researcher usually has little control over the 'when' and 'to whom' the treatment (exposure, intervention) is directed.

= **No Exposure [=treatment] Randomisation**

This is often true in the medical situation. A similar situation is when you have two groups and the independent variable you are interested in investigating is something like expertise, say novice and expert footballers, you can not randomly assign subjects to each group because it is predetermined.

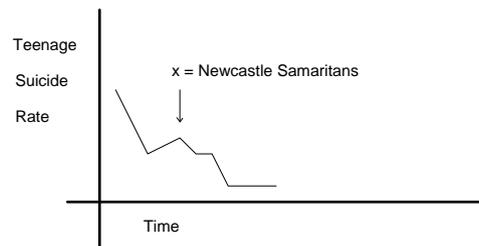
Question: How does the above treatment randomisation differ from Sampling Randomisation?

### 4.2.1 Interrupted Time-Series Design

In this design a group is followed over time.

M = Measure

X = Treatment / intervention (proposed?)



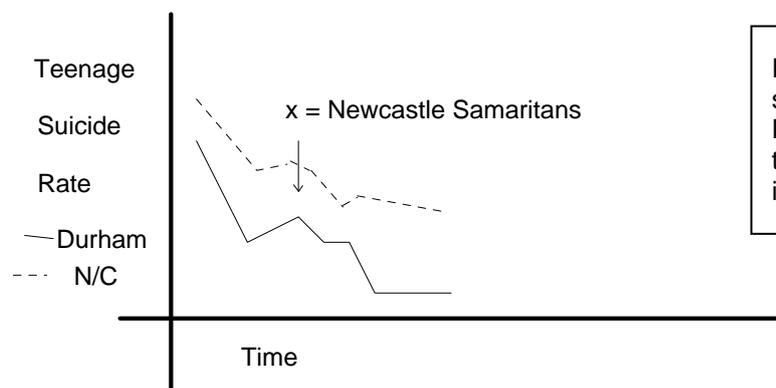
There is a problem here - did the treatment cause the effect or was it something else. This is an example of the saying 'correlation does not mean causation'.

An example is the Longitudinal study by the OPCS which constantly monitors a number of people born at a specific time (Cohort).

### 4.2.2 Multiple Time-Series Designs

Note: 'Multiple' refers to the number of groups not the number of measurements.

This design allows comparison, even possibly with another group that can act as some sort of control but WITHOUT RANDOMISATION. Therefore it provides some control over 'history' (that is other factors effecting measure during time span of investigation other than the proposed variable)



For example consider monitoring the standardised suicide rate in both Durham and Newcastle, During which time a new Samaritans centre is set up in Newcastle.

### 4.2.3 Non-equivalent Control Group Design

This is a very common widespread field design.

	Time 1	Time 2	Time 3
Non-randomised group 1	Pretest	Experimental Treatment	Posttest
Non-randomised group 2	Pretest	No Treatment	Posttest

Although the groups have not been randomised looking at all the important (?) pre-test characteristics afterwards may show a high degree of similarity (i.e. drawn from the same population). If so, this usually indicates a high degree of internal validity particularly in terms of 'history' and 'maturation'.

### 4.2.4 Separate-Sample Pretest-Posttest Design

In this design **randomisation** into each of the groups is used.

Group	Time 1	Time 2	Time 3
Randomised group1	Pretest	Probably No Experimental Treatment	No Posttest
Randomised group2	No Pretest	Experimental Treatment	Posttest

Is is a good design for large studies. However, there is danger of contamination of the treatment in group 1. The design results in a high degree of external validity but a low degree of internal validity.

To solve these problems the following two groups are added to the design:

### 4.2.5 Separate Sample Pretest-Posttest Control Group Design

randomised group3	Pretest	No Treatment	No Posttest
randomised group4	No Pretest	No Treatment (Controls	Posttest

The above four group design is the only one that offers a high degree of internal validity and only has the threats to external validity[generalisation] of Multiple treatment interference and poor Operationalisation.

#### Exercise:

Compare the above with the Solomon design?

### 4.3 Non-experimental Designs

In these designs it is only strictly possible to investigate associations, the direction of any causal link is purely due to the researchers interpretation of the results. However it is frequently a matter of 'common sense' to suggest a direction for any correlation's, if one does exist.

#### 4.3.1 One-Shot [Cross-sectional] Study [Case Study][Survey]

In this design one measurement is taken, possibly after some type of intervention.

time 1 - Treatment?

time 2 - Posttest

Why is this design so bad? Consider the scenario where an investigator measures the effectiveness of psychoanalysis on patient recovery by giving a questionnaire to patients having completed the course. By representing this situation in table form the problem can clearly be seen:

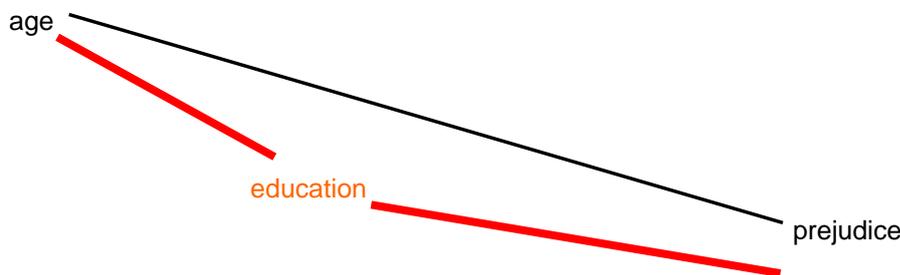
		Independent variable (Treatment)	
		Treatment group	Non Treatment (or alternative treat.) group
Dependent Variable (Measure)	Success	 Cross-sectional study	
	Failure	Recidivism?	
		Comparison	

In effect we are only looking at one of four possible groups of subjects, we have no representatives from any alternative or control group. Significantly, we also have no data from patients that may have dropped out during the course of the trial.

#### 4.3.2 One-Shot Correlational Study

This is the process of discovering relationships between two or more variables by analysing surveys. This method made up 60% of Social sciences research<sup>5</sup> in 1975 and I bet it is even higher in 2005?

Drawing conclusions (inferences) from correlations obtained in a one-shot survey can be **Very dangerous**. For example, by analysing survey results you might discover a correlation between age and some type of prejudice and also find a correlation between education and prejudice along with age and education. You are presented with the quandary of which causes which? Such a situation is referred to as **confounding [of the dependent variable]**.



#### Exercise:

Why is analysing data afterwards ('ex post facto') to try to find relationships not ideal?

<sup>5</sup>Smith H W. 1975 Strategies of Social Research. p101 London Prentice Hall International.

### 4.3.3 One-Group Pretest-Posttest Design

This design should only be used when no other available.

Time 1 - Pretest

time 2 - Treatment?

time 3 - Posttest

### 4.3.4 Static-Group Comparison Design

This design is the same as the 'post-test only control group design' except there is no control group.

Time 1	Time 2	Time 3
-	Experimental Treatment	Posttest
-	No Treatment	Posttest

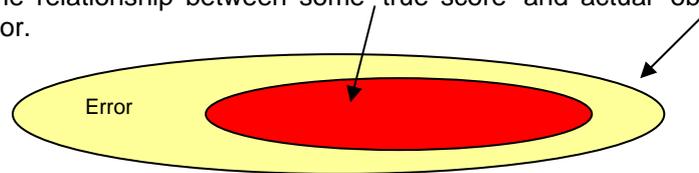
### Exercises

- 1 Draw a tree diagram of the designs discussed in this chapter, detailing their similarities and differences.
- 2 Design a classic experiment.
- 3 Design a Solomon four group experiment.
- 4 Design a Multiple time series quasi experiment using two groups.

## 5. Reliability

**A definition:** Level of consistency of a particular measure / procedure<sup>6</sup>.

It can be considered to be the relationship between some 'true score' and actual 'observed scores' which includes a degree of error.



All measures have an acceptance [tolerance] level (e.g. number of false results or deviation from 'true score' which is acceptable)

Influences on reliability of a measure include:

- Sample size
- Number of equivalent measures (more the better up to 8)
- Instrument (including human) characteristics

### 5.1 Testing Reliability

This depends partly upon the particular instrument used. Some ways it can be assessed are by:

1. **Test-Retest** - Same sample tested twice. Time interval between assumes 'Stability'. (I.e. not History effects in internal validity)
2. **Equivalent [alternative/multiple/pure] forms [measures]** - use different measures for same thing.
3. **Cronbachs  $\alpha$  correlation** - statistical test to show degree of similarity between different measures. This enables you to develop equivalent measures
4. **Split half random allocation** - used in questionnaires two different sections measure the same thing. Compare results between the two sections.
5. **Inter-rater reliability** - used to assess the reliability between different scorers (e.g. two doctors assessing same subject for degree of neurosis).

The problem of inter-rater reliability is an ongoing one. For an article in the Accident and Emergency area see Zoltie N Dombal FT de 1993 The hit and miss of Injury severity score (ISS) and combined Trauma and Injury Severity Score (TRISS). BMJ 307 906 - 9 (October 9th). This article also demonstrates how to do it along with the standard references for this area of research.

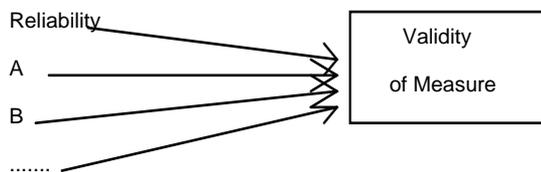
It is interesting to realise that Sir France Galton one of the first psychologists said this when addressing mental:

"There are many faculties that may be said to be potentially constant in adults though they are not developed, owing to want of exercise. After adequate practice, a limit of efficiency would in each case be attained and this would be the personal constant (*italics added*); but it is obviously impossible to guess what that constant would be from the results of a single trial. No test professes to do more than show the efficiency of the faculty at the time it was applied, and many tests do even less than this" (Galton (1885), in Pearson, Vol. II, pp. 371-2).

### 5.2 Validity (part 1)

How is validity of a measure related to its reliability? It is very difficult to provide a definition of what Validity means and we will see there are many different types of it. For now think of validity as being the degree to which something measures what you actually think it is measuring. For example imagine that a research project is looking at depression in teachers and for the measure of depression the researchers decided to count the number of days off work within the last year as being an indicator. Is this a valid piece of research?

<sup>6</sup>For a mathematical perspective, see Marrow J R. Jackson A W. 1993? How significant is your reliability. 64 3 352-355. Provides tables for the significance of a reliability measure.



Because the validity of the measure is dependent upon how reliable the particular measure is it is always less than the reliability.

**Exercise:**

What other factors do you think might affect the validity of a measure?

## 6. Operationalisation

Operationalisation is the process of converting a concept into a measure. This measure can be a variable, constant or scale depending upon the situation. We will consider several examples, some good, and some bad.

### 6.1 Concepts

Concepts are the 'things' that we are trying to measure in research. Here are a few examples:

- Motivation
- Height
- Self perception
- Aggression
- IQ
- Weight
- Exhaustion
- Skill
- Health
- Heart Rate
- Performance level
- Ability
- Auditory hallucinations
- Learning disability

**Exercise:** From the above list mark alongside each item how you think you might measure each in a piece of research.

Concepts are often divided up into those that can be measured 'directly' and those that can not be. Those that can only be measured by their 'shadows' are referred to as '**Constructs**'. Using this scheme Height is measured easily, but motivation can only be measured by a possible set of different measures.

#### 6.1.1 Dimensions of a Concept

In the above examples, the concept height only requires one measure to adequately describe it. However, IQ requires a whole set of measures relating to particular aspects of it:

- Memory
- Motor Skills?
- Comprehension
- Analysis
- etc.

Each of the above aspects can be considered to be different **factors** or **dimensions** of the concept. We could say therefore that IQ is **multidimensional** (i.e. having more than one dimension).

**Exercise:** Consider the information you collected for the previous exercise. Choose those concepts that you thought required more than one measure to provide an adequate definition and attempt to classify the measures into groups.

## 6.2 Measures

A measure can be a variable, constant or scale depending upon the situation.

### 6.2.1 Variables & Constants

A constant is a characteristic that has the same value over the entire sample (it does not vary). A variable is something that can take more than one value. Variables can sometimes be subdivided into dependent and independent variables.

A **dependent** variable also called the **effect**, **response** or **outcome** variable is what the researcher measures before and / or after manipulating the **independent** variable (also called the **treatment** or **intervention** variable). Often the independent variable is of nominal measurement type. The researcher may **actively manipulate** the independent variable and see what effect this has upon the dependent variable or, **passively observe** thereby collecting a series of measurements. In the second situation, she will have to work out the direction of the relationship by investigating which variable changes first. To do this she will investigate the **lag/response time** between a change in the independent variable to the observed change in the dependent variable. For example to observe the effect of amphetamines (i.e. independent variable) upon running performance (i.e. dependent variable) it would be necessary to wait until the drug had been absorbed through the stomach.

Often the term 'measure' is used to refer to the dependent variable. I will use the term to include both independent and dependent variables.

**Example 1** Active manipulation = assigning subjects to several groups each receiving a different regime.

**Example 2** Passive observation = observing over a series of years the effect different law enforcement strategies have upon road traffic accidents.

**Example 3** A researcher investigating the effects of two different types of training programmes on half marathon performance. She randomly assigns half her subjects to one type of training and the rest to the other. She measures performance pre and post training programme for all subjects.

Independent variable (factor) = Training type = nominal measurement ( two levels)

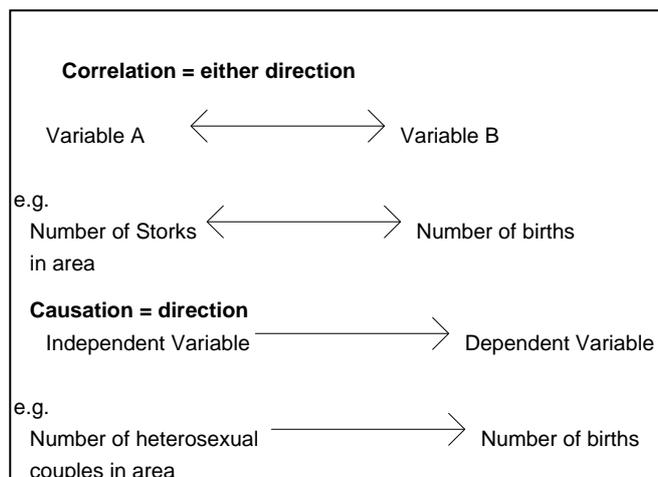
Dependent variable ('measure') = performance = measurement?

#### Exercise:

Consider the advantages and disadvantages of the active and passive manipulation of the independent variable.

In other words, the value the dependent variable takes depends upon the independent variable.

This is a **causal relationship and is different from a correlation.**



**Exercise:**

Think up several relationships in sport or psychology and decide if there is a causal or just a correlational relationship.

## 6.2.2 Scales

These are also called **indexes**. This is a measure that consists of two or more items. These items are themselves measures. The best way to understand this is to give a few examples:

Example 1 Socio-economic status = Income + Education + ?

Example 2 Activities of daily living scale =

- Eating
- Dressing / undressing
- Care for own appearance
- Walking
- In / out of bed
- Bath / shower
- Get to toilet
- Done above tasks for 6 months

For each item a score of 1 is assigned if the patient was unable to do the activity, 2 if the patient was able to do the activity with assistance, and 3 if the patient required no assistance.

Example 3 Motivation of runners scale (MOMS).

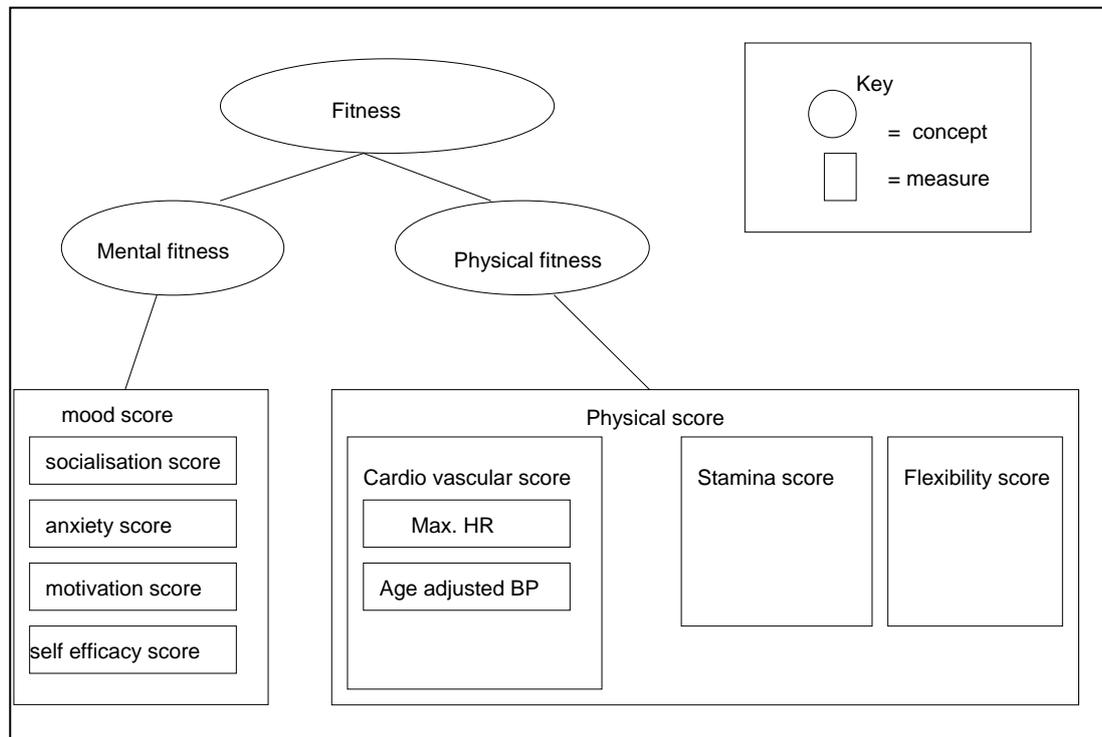
See Masters KS Ogles MB Jolton JA 1993 The development of an instrument to measure motivation for marathon running. Research Quarterly for exercise and sport 64 (2) 134 - 143

**Exercise:** Suggest a possible set of items that could be included in an index measuring a topic area you are currently studying.

### 6.3 Graphical Method of Representing Operationalisation

Because operationalisation is so important and how well it has been done often provides the litmus test for a piece of research. I suggest you get into the habit when reading research articles of using the following method to analyse the various concepts and measures presented.

I will consider one example. Fitness:



Things to note:

The ovals represent concepts, the researcher may well consider several concepts and decide upon measuring a particular smaller concept. For example in the above a researcher may decide only to concentrate on physical fitness rather than mental fitness. This is fine as long as the researcher is aware of the limitations. It would be misleading to title an article fitness evaluation if s/he did take this approach.

The squares represent measures, where groups ('batteries') of measure provide a score of some sort I have placed them in another box. Clearly, this is just my way of helping me understand the situation.

**Exercise:** The above example is incomplete and left for you as an exercise to complete.

**Exercise:** Take an article from one of the sports or medical journals and consider the various concepts and measures used. Use the above technique to analyse the article. Do not be too strict with yourself concerning the diagram details.

## 7. Validity (part 2)

We can now provide a little more clarity in our definition of validity. We can now say that validity of a measure is the degree to which the measure or measure equals the concept which it / they are attempting to measure.

Often researchers use well validated measures rather than trying to attempt to develop their own. Besides the obvious advantage that this is a valid measure two other advantages are:

- Less effort is required to repeat an experiment
- Results can be compared with previous experiments.

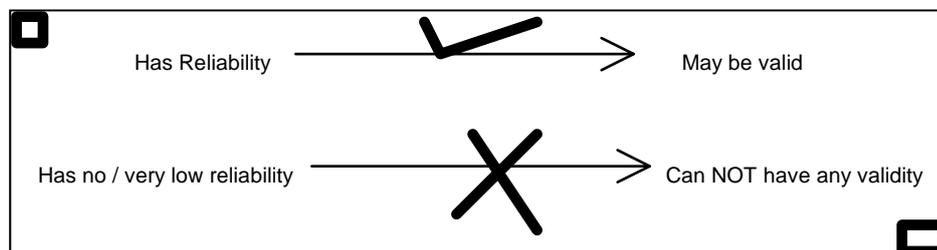
### 7.1.1 Relationship with Reliability

Reliability is the degree of consistency of a measure or degree of error of a measure. That is if all else stays the same we would expect to obtain the same value repeatedly from the measure if it were reliable.

It should be noted that a measure can be reliable but not necessarily valid. A researcher may believe s/he is measuring perceived exertion by measuring heart rate with a heart rate monitor. Most likely the heart rate monitor provides a very accurate measure of heart rate but it may not reflect the level of perceived exertion the subject is feeling. You can therefore have a measure that is very reliable but totally invalid. Lets now consider the opposite situation.

Can we have a measure that is valid but not very reliable? A valid measurement implies a relatively accurate measure (Portney & Watkins 1994 p.69). For example, if you decide to measure heart rate by palpating the radial artery and using a watch with only a minute hand for timing you are unlikely to produce accurate heart rate measurements. You may assume you are measuring heart rate but most of the time the results will not reflect this concept.

Therefore:



### 7.1.2 Types of Validity

There are several types of validity, only one is 'validity of measurements' Further details can be found in (Robson 1993, Portney & Watkins 1994, Graziano & Raulin 1993, Cook & Campbell 1979). Details of these other types of validity are given in subsequent chapters.. We will consider a little about validity of measurements in this chapter Besides the references given above one article to start with is Taylor J A review of Validity Issues in Sport Psychological Research. Journal of Sport Behaviour 10, 1 3-12. Be warned this is a difficult read and uses different terminology to me in several instances.

We will discuss just three types of validity of measurements; Face (logical), Content and construct .

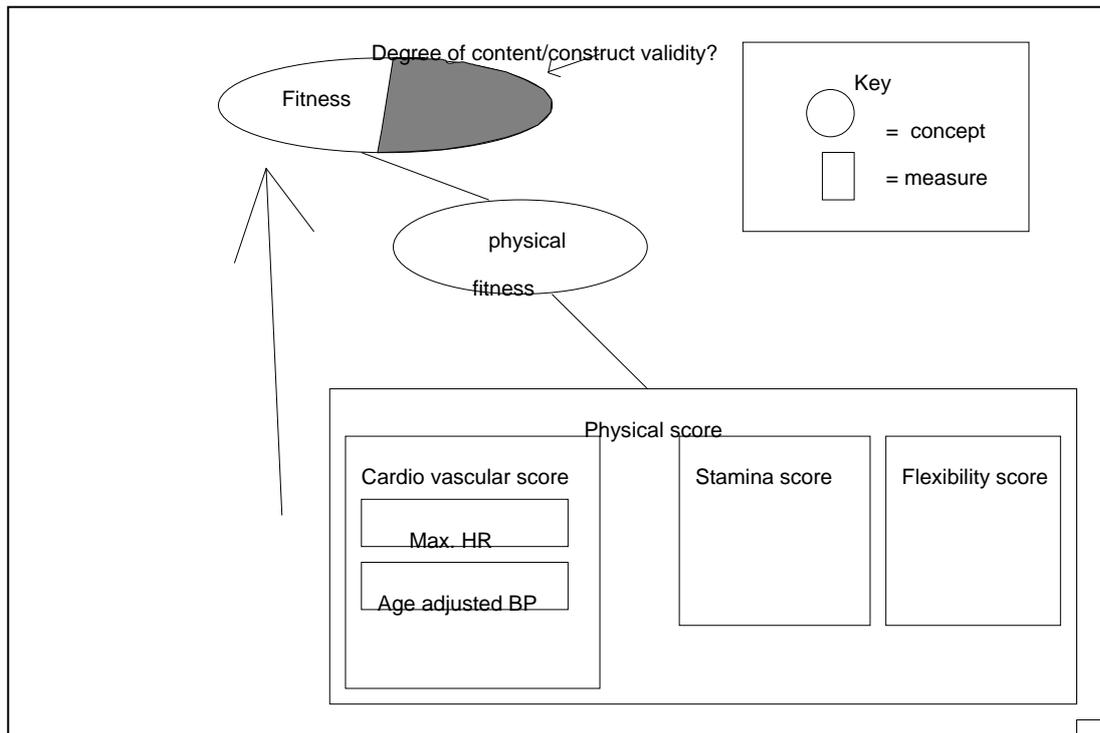
### 7.1.3 Face (logical), Content and Construct Validity of Measurements

**Face (logical)** validity is the degree of belief that the researcher has that the measure appears to measure what it is supposed to. It is therefore often called a qualitative measure.

**Construct validity** is the degree to which the constructs of interest have been effectively operationalised (Cook & Campbell 1979). **Content validity** is the same thing but this time it can be either a construct or a concept. This is, at least, how I interpret the situation between construct and content validity, I always feel a little uneasy about this, if you think I'm wrong I would be quite happy to discuss it with you. Cook and Campbell and Robson 1993 do not mention content validity (see Cook & Campbell 1979. p.38 and p.59). Portney and Watkins 1993 p.71-77 provides a detailed description of each type. Oppenheim 1992 p.162 considers content validity to be concerned with how well balanced

the measures are in relationship to the 'content domain' to be measured. But fails to define content domain.

I believe both content and construct validity can be expressed using, our diagram developed to show the process of operationalisation, except this time we work backwards.



From the above you will have begun to realise that in reading the literature you will find many differing and conflicting definitions of validity. This is particularly the case concerning the various types of validity of measurements, If you want to investigate this it is best to go back to the source i.e. Cook & Campbell 1979.

## 7.2 Importance of Good Operationalisation

The importance of carrying out this process correctly cannot be underestimated. Clearly, a whole research project may be a total waste of time if it is not measuring what it is supposed to. This process is equally important for the independent and dependent variables.

**Exercise:** By considering some of the variables defined in examples given earlier in this chapter decide the degree of validity of measurements they each possess

## 7.3 Context Specific

A context is another word for situation. Some measures work well in all situations (contexts) e.g. a ruler for measuring height. However, some only work well in specific situations e.g. a mercury thermometer would not work on the moon? Because of this, the validity of a measure is said to be context specific. Obviously part of the context consists of the researcher(s) and the subjects. This suggests that a measure developed using an all female group of subjects may not necessarily be equally as valid when used on a group of males.

## 7.4 Summary

Because the above section contains a large amount of material I have included a summary at this point.

The above section has mainly been concerned with the process of operationalisation. However, to understand this process fully it was necessary to discuss what concepts, constructs and measures are along with 'validity of measurements'. Various types of variables and causal / correlational relationships were also presented.

**Exercise:** Draw a concept map of the above section.

## 8. Hypotheses

Most courses concerned with research present hypotheses right at the beginning suggesting that the first thing anyone must do is to define a hypothesis of one sort or another in research. This is very commendable and certainly the correct way for researchers to work. However often to the uninitiated the production of a hypothesis is one of the most difficult of tasks. Because of this I have decided to tackle this subject near the end of the course. We will consider each hypothesis type in turn and then see how they relate to each other.

### 8.1 The Research Hypothesis

This is what most people imagine when you ask them what a hypothesis is. As usual different writers have different views as to what it actually is. Taking one of the most laze views Robson 1993 p.28, describes it as a statement which is a "tentative guess, or intuitive hunch". Smith, 1975 P.40 does not talk about a research hypothesis but in the same context refers to a **scientific hypothesis** which he defines as:

"[a statement] consisting of two or more variables linked by some relationship(s)"

In this instance we therefore must have operationalised our concepts before we can define our hypothesis.

In the rest of this section I will work using Smiths 1975 definition of a research/scientific hypothesis.

All the above has been rather abstract. We will now consider some examples.

**Example 1** The prevalence of heart disease is associated with cigarette smoking

Variables = heart disease, cigarette smoking

Relationship = association

**Example 2** Footballers perform better the more motivated they are

Variables = Football players, Motivation

Relationship = relationship

**Example 3** There will be a increase in swimming performance when undertaking a new training method.

Variables = swimming performance, training method

Relationship = increase

**Example 4** Novice and experienced weight lifters produce the same amounts of lactic acid when training

Variables = Weight lifters experience level (novice, experienced), Training, lactic acid production

Relationship = no difference

**Example 5** A course of anabolic steroids will increase running performance

Variables = running performance

Relationship = increase

**Example 6** A course of pre performance focusing sessions will reduce performance anxiety

Variables = focusing session, performance anxiety

Relationship = reduction

**Exercise:** Make up 10 of your own.

It would appear from the above that you could possibly divide up the relationships into two main types:

Correlation's/causation's      and      Differences

## 8.2 The Statistical Hypotheses

This is the next level of complexity. Here is a first stab at what it actually is:

### "Statistical Hypotheses

This is a particular statement concerning the value of a characteristic of a population. It is therefore usually stating an inference from a sample which we know about to something which is unknowable, the population. There are two types: Simple and Composite. Simple statistical hypotheses are those that have one unique value. Composite statistical hypotheses are those that do not have a single unique value. For example:

Population average for male sports students strength = 12  
= a simple statistical hypothesis

Population average for male sports students strength does NOT =12  
= a composite statistical hypothesis "

What additional characteristics has a statistical hypothesis got? Firstly it refers to a population. Often this population is called the **target population**. This means a population defined by the researcher undertaking the study. For example, a researcher may be investigating the use of daily exercise regimes in a local home for the elderly. Her target population might be all elderly people or just those in the particular type of home she is working at. It would be her responsibility to state which was her target population along with reasons for her decision.

Secondly, we now have to state either one value or a set of them. To show how this works in practice we will once again consider the examples given above:

**Example 1** The prevalence of heart disease is associated with cigarette smoking.

Simple statistical hypothesis = There is no relationship between cigarette smoking and heart disease  
e.g. estimated population correlation = 0

Composite statistical hypothesis = There is a relationship between cigarette smoking and heart disease  
e.g. estimated population correlation  $\neq$  0

**Example 2** Footballers perform better the more motivated they are.

Simple statistical hypothesis = There is no relationship between performance and motivation in footballers. E.g. estimated population correlation = 0

Composite statistical hypothesis = There is a relationship between performance and motivation in footballers. E.g. estimated population correlation  $\neq$  0

**Example 3** There will be an increase in swimming performance when undertaking a new training method.

Simple statistical hypothesis = There is no difference in swimming performance when undertaking a new training method. Note: This is a simple statistical hypothesis because 'no difference' = 0 i.e. one value.

Composite statistical hypothesis = There is a difference in swimming performance when undertaking a new training method. Note: This is a composite statistical hypothesis because 'a difference' = more than one possible value

**Example 4** Novice and experienced weight lifters produce the same amounts of lactic acid when training

Simple statistical hypothesis = There is no difference between novice and experienced weight lifters in the production of lactic acid. Note: This is a simple statistical hypothesis because 'no difference' = 0 i.e. one value.

Composite statistical hypothesis = There is a difference between novice and experienced weight lifters in the production of lactic acid. Note: This is a composite statistical hypothesis because 'a difference' = more than one possible value

**Example 5** A course of anabolic steroids will increase running performance

Simple statistical hypothesis = There is no difference between running performance if you take anabolic steroids or not. Note: This is a simple statistical hypothesis because 'no difference' = 0 i.e. one value.

Composite statistical hypothesis = There is a difference (an improvement) between running performance if you take anabolic steroids or not. Note: This is a composite statistical hypothesis because 'a difference' = more than one possible value

**Example 6** A course of pre performance focusing sessions will reduce performance anxiety

Simple statistical hypothesis = There is no difference between anxiety levels if you carry out pre-performance focusing sessions or not. Note: This is a simple statistical hypothesis because 'no difference' = 0 i.e. one value.

Composite statistical hypothesis = There is a difference (a reduction) between anxiety levels if you carry out pre-performance focusing sessions or not Note: This is a composite statistical hypothesis because 'a difference' = more than one possible value

**Exercise:** Carry out the same procedure for the 10 research hypotheses of your own you developed earlier.

### 8.2.1 Null Hypotheses( $H_0$ )

Given below is a first stab at defining a null hypothesis:

"The null hypothesis is a particular type of simple statistical hypothesis which has been developed by statisticians. The essential characteristic is that it can be evaluated in terms of a probability given certain assumptions are met.

e.g.  $p(\text{obtaining a sample of ten sports students with an average strength rating of 10. given that the population average is 12}) = 0.75$ "

The above definition is vague because you need to be aware of, and understand Probability Density Functions, inferential statistics, sampling distributions, and the 'statistical test' to fully appreciate what a null hypothesis is. These aspects should have been covered in your introductory statistics course.

**Exercise:** What are the components of a statistical test?

From elementary probability theory the above statement can be rewritten as a conditional probability  $p(\text{sample data mean} = 10 \mid \text{mean} = 12 \text{ in population}) = 0.75$  Remember that the '|' symbol is interpreted as 'given that'.

The probability is therefore dependent upon assuming a particular value for a population characteristic.

#### 8.2.1.1 Testable Hypotheses

The one most important aspect of the null hypothesis is that it is testable (i.e produces a probability given a specific condition). Furthermore, the probability is valid if you keep to the relevant data assumptions for the particular inferential statistic being considered.

Once again, using the same examples we will now develop the null hypothesis for each of them.

**Example 1** The prevalence of heart disease is associated with cigarette smoking

Null hypothesis =  $p(\text{heart disease, cigarette smoking data} \mid \text{correlation} = 0 \text{ in population})$

**Example 2** Footballers perform better the more motivated they are.

Null hypothesis =  $p(\text{performance, motivation} \mid \text{correlation} = 0 \text{ in population})$

**Example 3** There will be a increase (= difference) in swimming performance (=sp) when undertaking a new training method.

Null hypothesis =  $p(\text{training old sp, training new sp} \mid \text{difference} = 0 \text{ in population})$

**Example 4** Novice and experienced weight lifters produce the same amounts of lactic acid (=la) when training

Null hypothesis =  $p(\text{novice la, experienced la} \mid \text{difference} = 0 \text{ in population})$

**Example 5** A course of anabolic steroids will increase (=difference in) running performance (rp)

Null hypothesis =  $p(\text{no steroids rp, with steroids rp} \mid \text{difference} = 0 \text{ in population})$

**Example 6** A course of pre performance focusing sessions (fs) will reduce performance anxiety

Null hypothesis =  $p(\text{no fs performance anxiety, with fs performance anxiety} \mid \text{difference} = 0 \text{ in population})$

**Exercise:** Carry out the same procedure for the 10 research hypotheses of your own you developed earlier.

### **8.2.1.2 Why 'Null'**

This is a question students often ask, there are numerous reasons. I'm never quite sure how many of them are due to the fact that everyone tries to explain it away. Anyway, here are some of the explanations:

Because the estimated population parameter is usually 0

Because we would like to reject = NULLIFY it.

Because there is no difference between the estimated population values obtained from samples for a particular characteristic (e.g. means with the t statistic) we are measuring with the statistic.

The complete Oxford English dictionary provides the historical development of the term null hypothesis, and is quite interesting to read.

### 8.2.2 Relationship between a Inferential Statistic and a Null Hypothesis

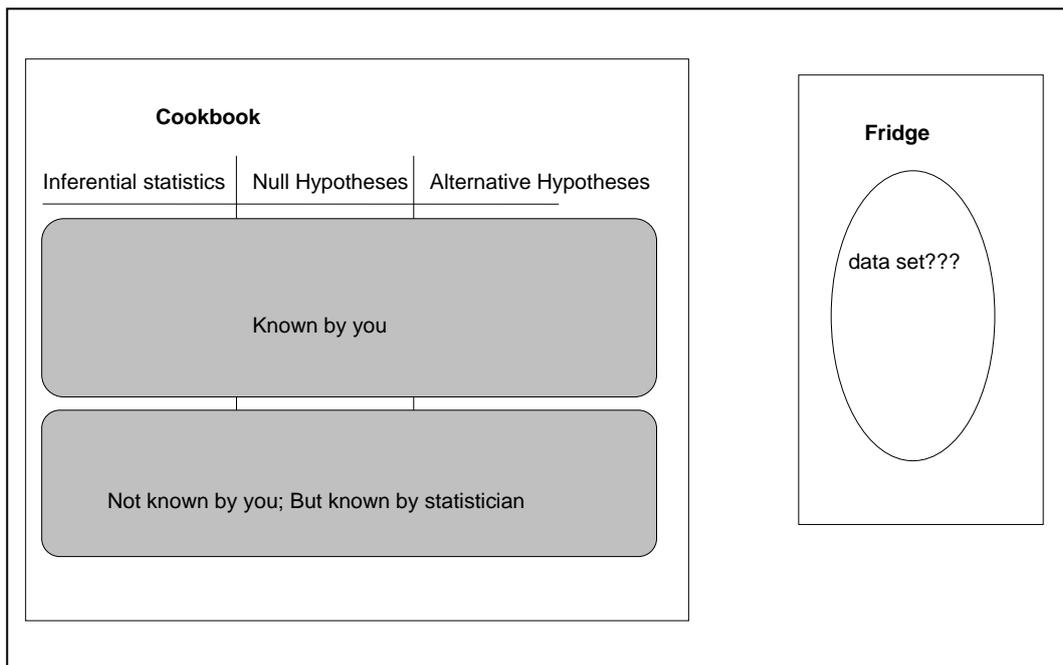
Just as we interpret the 'p' probability statements into sentences - narratives. A particular null hypothesis can be considered the narrative interpretation of a particular inferential statistic. So for every inferential statistic there exists a null hypothesis. The table below provides some examples of common inferential statistics:

Inferential statistic	Null Hypothesis
Paired t statistic	Estimated Population mean = 0 for difference scores
independent sample t statistic	Estimated Population mean is same for both groups (therefore est. mean $\mu_1$ - est. mean $\mu_2 = 0$ hence 'null')
Wilcoxon matched pairs statistic	Both paired samples come from identical populations (difference scores symmetric about 0 hence 'null')
Mann Whitney U	Both independent samples come from identical populations (sum of ranks for both groups symmetric about 0 hence 'null')

The above relationship explains the difficulty students, new to research methods and having no understanding of inferential statistics, have when asked to devise null hypotheses.

The situation can be considered similar to:

Someone possessing a collection of food in the kitchen (the data) and expecting them to have memorised an extremely large number of tried and tested recipes (null hypotheses) when they have never used a cookbook (inferential statistics) before.

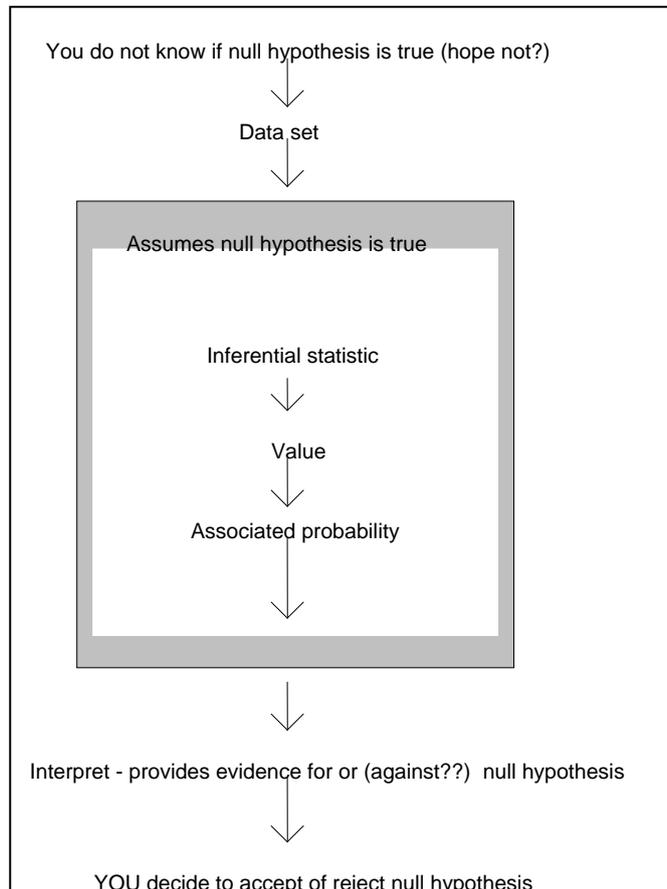


Because the null and alternative hypotheses are narrative interpretations of the inferential statistic they are not so precise, hence you will find different books providing different definitions. Howell 1993 p.617 provides an example of how a null hypothesis can be increasingly accurate with regards to the underlying inferential statistic.

Just to labour the point:

A null hypothesis is a narrative interpretation of a specific inferential statistic

Furthermore, because a null hypothesis is a narrative interpretation of an inferential statistic it is equivalent to saying that the probability obtained is that 'under the null hypothesis being true'. In fact the inferential statistic has only one reality, precisely that, of  $h_0$  being true. WE DECIDE TO REJECT OR ACCEPT THE NULL HYPOTHESIS not the inferential statistic.

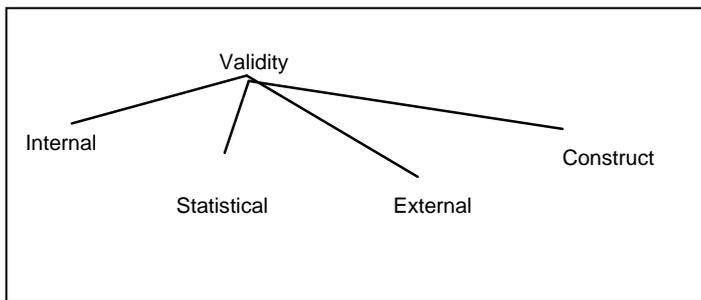


### 8.2.3 Alternative Hypothesis ( $h_1$ )

This is the logical opposite to the null hypothesis. For example, if the null hypothesis states there is no relationship between height and weight the alternative hypothesis states that there is. Similarly, if the null hypothesis states that the estimated population means for two groups is the same the alternative hypothesis states that they are not.

The alternative hypothesis is therefore usually composite and is the equivalent to the research or scientific hypothesis which we discussed at the start of this session.

## 9. Validity (part 3)



This is the third time we have visited this topic and we will now look initially at the various types of validity based on Cook & Campbell 1979 who considered the 4 types in the diagram (p.39).

### 9.1 Internal Validity (of Findings)

Internal validity is concerned with the components of the individual research design rather than drawing inferences concerning a wider population. Basically, you are asking here:

'did my treatment really have that effect or was it something else'

We are looking at the level of individual subjects and specific groups (i.e. control and treatment) within the sample. NOT the entire sample. We are particularly interested in how particular subjects or more importantly, groups may be effected by something other than the independent variable (= Confounding variables). The reason for this is that if the thing affects all the groups and therefore subjects equally it will not change our results. The gap between the control and treatment groups will be the same as if it did not exist.

Campbell, 1957 and Cook & Campbell 1979 has clearly defined factors that have a **negative influence** ('threats') on internal validity. These 'classes of confounding variables' usually arise for two reasons: through groups in the design being non - equivalent at the start of the research and researchers treating them differently during the research period. These threats are:

1. **History** - This is any factor external to those being measured that may change **during the research** that affects the dependent variable to varying degrees between groups (e.g. war / employment / cost of living). If it affects both the control and treatment group(s) equally there is no problem (see chapter controlling the treats to the various types of validity). If there is no control group the researcher can not be sure if the change in the dependent variable is due to the treatment or some 'history' effect.
2. **Testing** - Individual subject's reactions to testing during the research. e.g. Masters & Johnsons experiments.[**Reactive effects to testing**].
3. **Subject Maturation** - e.g. effects of age on subject when repeating measurements.
4. **Instrument / Researcher Maturation** - e.g. Equipment wearing out, interviewers getting better/tired. etc.
5. **Selection Bias** - e.g. non-random selection of subjects as in convenience samples (e.g. volunteers / prisoners etc.) means that individual groups may react differently to the treatment, a certain type of individual volunteering for a specific group.
6. **Mortality Bias** - e.g. different groups in an experiment may have different drop out rates not due to the treatment ( leaving area rather than dying). Groups that are initially equivalent may not be at the end of the trial due to this effect.
7. **Equilibrating [regression] effects** - This is a peculiar effect when carrying out repeated measures. and can best be explained by way of examples viz.:

If you measure someone who is excessively tall or short the next time you measure them it is more likely that their measurement will be nearer the mean. This phenomenon also occurs over time. Fore example, excessively tall / short parents produce children of more average height leading to the incorrect conclusion that the population is losing its height diversity. Similarly, if you make a measurement of mean height next time the measurement will be further from the mean (similarly parents of mean height are more likely to have children of height deviating more from the mean).

8. **Selection-Maturation Interaction Effects** - Sample selection often works in conjunction with maturation, history, testing etc. to produce spurious results. e.g. a researcher comparing two groups of student volunteers where one group, unknown to him/her, have a past history of taking aptitude tests. This 'history selection bias' combination might differentially effect each groups answers to attitude scales/tests.
9. **Ambiguity about the direction of a causal relationship** - Does A cause B or B cause A or are they just correlated.
10. **Diffusion or Imitation of treatments** - If the groups can communicate with each other they may share / swap the treatment etc. Done by subject
11. **Compensatory Equalization of treatments** - This is when one group receives special treatment. For example, in a new training regime various pressures may be placed upon the researchers carrying out the testing which may result in the control group having some other hidden treatment. For example, the trainer may feel sorry for the control group and offer additional unrecorded sessions. This is not the result of something the subject does but something the researcher does because they are aware of the groups differences.
12. **Compensatory Rivalry** - Same as above but done by subjects themselves. Control group subjects deliberately doing hidden sessions to keep up with those undergoing the new training.
13. **Resentful demoralization of subjects receiving less desirable treatments** - The control or placebo group may be aware they are not getting the 'best' treatment. They may become demoralised and give up. Comparing there results with that of the treatment group will be different not necessarily because of the treatment effect but because of the negative effect of the placebo / control.

**Exercise:** Draw a diagram of the above situations. Start by including the researcher, subject and instrument.

## 9.2 Validity of Statistics

This is concerned with the appropriateness of the statistical tests. It is divided up into:

1. Violated assumptions of statistics
2. Low statistical power
3. Error rate
4. Reliability
5. Variance

**Violated assumptions of statistics** - Most inferential statistics make a number of sample data assumptions. These should always be checked before carrying out the inferential statistic. The most common one is that the values are normally distributed.

Also remember visual assumptions of graphs, people expect to see the entire axis, plotting means instead of raw data where appropriate. There are a large number of possible visual mis-representations etc.

**Low statistical power** - Power is determined by sample size, treatment effect, reliability, variance and the particular inferential statistic used. It is defined as the probability of correctly rejecting a null hypothesis when in reality it is false. We therefore want as much power in the research design as possible. If the researcher has managed to reject the null hypothesis at a sensible critical value power is not an issue. Power needs to be considered at two stages in the research design.

Firstly, it dictates the sample size necessary to ensure that the researcher actually has a probability of being able to reject the null hypothesis. A research design with zero power would mean that whatever the result of the research the researcher would be unable to reject the null hypothesis - It would therefore be a waste of time and money carrying it out.

Secondly, a power analysis is indicated when the researcher has obtained an insignificant result (i.e. been unable to reject the null hypothesis).

The easiest introductory text that I have come upon explaining power issues is chapter 6 'What does not significant really mean' in Primer of Biostatistics by Stanton A Glantz 3rd ed. 1992 McGraw-Hill.

**Error rate** - If you carry on doing enough inferential statistics you are bound to find a significant result by chance in the end. If multiple inferential statistics are carried out it is essential to use procedures to prevent this happening.

**Reliability and Variance** - If the instrument is unreliable it can not produce valid data. Examples are uncalibrated machinery, unstandardised questionnaire delivery, poor protocols, basically any type of poorly standardised testing procedure. This will result in results with a large error variance or 'noise'. The timing of the measure is also of critical importance particularly when measuring 'peaks' such as maximum heart rate or lactic acid etc.

An alternative way of approaching the above threats to internal validity is to consider extraneous variables as being due to:

1. External factors - those factors that emerge from the environment and the experimental situation.
2. Intrinsic factors - those that emerge from personal characteristics of the research components.

(After Portney & Watkins p.125 adapted and extended).

In contrast to the above two sections the next two are concerned with external validity, that is generalising the results to a wider sample/population, context or time.

We are no longer interested in what might affect one group of subjects differently to another in a research design other than the independent variable. Now we are interested in what might have affected the ENTIRE sample over all the groups studied in the research. In other words can we generalise our findings.

### 9.3 Construct Validity of Measurements

While this topic has been partially covered elsewhere we did not consider the threats to this type of validity. Cook & Campbell 1979 suggest 5 threats:

1. Poor Operational definitions of independent & dependent Variables
2. Time frame of operational definitions
3. Multiple treatment interactions
4. Experimental bias
5. Hawthorne effect

#### **Poor Operational definitions of independent & dependent Variables -**

**Inadequate definition of [outcome] dependent measure**<sup>7</sup> - Your dependent measure only provides information on your chosen measure not necessarily the concept which you believe it measures. In other words the results relate to the operationalised measure not to the underlying concept. If you have operationalised the concept badly the results will not necessarily relate to the concept. For example suppose you are measuring the concept of motivation. You operationalise the concept of motivation into a series of questions in a questionnaire. Unknown to you the questionnaire is actually measuring anxiety; you have a rather unique view of what motivation is. Because of this situation your results are valid for the measure but not the concept which you believed you were measuring. Furthermore the results are valid for your particular piece of research just that you have chosen a peculiar way of measuring motivation which unfortunately prevented you from comparing your results with other research concerning motivation, unless obviously they have also developed such strange ways of defining it.

**Inadequate definition of treatment**<sup>8</sup>[independent variable] - Same principle as above but this time concerning the definition of the treatment variable. In other words poor operationalisation of the independent variable this time. E.g. Treatment = 'the Robin' exercise regime which you operationalise as two one hour sessions a week, however the regime may actually entail a whole way of life. Once again your results will relate to your operational definition of the treatment rather than what it actually is as defined by 'the Robin' training institute ('the concept').

#### **Time frame of operational definitions**

A trial which follows the effect of a particular exercise regime for three months says nothing about what will be the effects in 6 months time. etc

#### **Multiple treatment interactions**

'Nothing is ever the same twice' e.g. 'pulsed' [repeated] treatments and crossover trials. Each treatment may have a long term effect. How often can someone do a marathon? It is not possible to infer results of multiple exposures from single exposure experiments and visa versa (e.g. does one drink have the same effect as drinking every day).

**Experimental bias** This is bias introduced into a study due to expectations of the subjects or the experimenter AS A WHOLE. Thus, the Hawthorne effect is one. Experimenters may present themselves in a particular way encouraging or discouraging a certain reaction to the entire sample. For an interesting account of the Hawthorne studies dispelling many of the myths see Waller 2002 (pp. 78 – 99).

**Hawthorn effect** This is the tendency that subjects have of performing better just because they have been paid attention (see Portney & Watkins 1993 p.141). NOTE: this refers to the entire sample not individuals or groups within it. This is just therefore the 'history' effect but affecting the entire sample equally.

---

<sup>7</sup>This is equivalent to Campbells 1957 'Irrelevant responsiveness of measures'.

<sup>8</sup>This is equivalent to Campbells 1957 'Irrelevant replicability of treatments'.

## 9.4 External Validity [Generalisations]

Cook & Campbell, 1979 has clearly defined factors that have a **negative influence** ('threats') on external validity. These problems arise from lack of naturalism and representation as well as the research design. I will use a modified version of the Cook and Campbell criteria based on LeCompte & Goetz 1982 referenced in Robson 1993:

1. **Sample [Selection] - Findings may be specific to sample studied e.g. volunteers, age, sex, occupation, sport?**

**Selection Bias Treatment Interaction Effects** - Biased samples (volunteers) may give unrepresentative responses to the independent variable(treatment) AS A WHOLE. e.g. It has been discovered that some researchers who carried out studies on marijuana effects using volunteers consisted of subjects over-representative of marijuana smokers, and were also into other drugs. The results would therefore be valid only for marijuana users rather than the general population. Paradoxically in this instance the results may therefore be more useful as the results would be more representative of people who were likely to be taking marijuana.

Some subjects (e.g. prison volunteers) learn to behave the way the researchers expect ('researcher expectation compliance') without the researchers realising it.

2. **Setting - Findings may be specific to the particular context in which the study took place, laboratory, field, university, sports club etc.**
3. **Time [history- second meaning] - The results may be due to the particular time the research took place. E.g. a new sports centre, pre second world war, pre NHS etc.**

We have now considered validity in three different sessions. The above should not be taken as gospel as there are numerous ways of considering validity and in the following section we will conclude our investigation of it by looking at how other writers classify validity.

## 9.5 Other types of Validity

Much of the interest in validity comes from educational research, for example one site divides it up into the following areas, similar to the way we did when we visited the concept the second time.:

- Content validity
  - Face validity
  - Curricular validity
- Criterion-related validity
  - Predictive validity
  - Concurrent validity
- Construct validity
  - Convergent validity
  - Discriminant validity
- Consequential validity

For details see

<http://professionals.collegeboard.com/higher-ed/validity/aces/handbook/evidence#contentvalid>

In the last few years the web has become a wonderful resource, and the following, from: <http://www.holah.karoo.net/validity.htm> written by Mark Holah, provides a clear succinct summary of both types of validity we have considered as well as some others:

There are three main ways of assessing the validity of a measuring tool:

(a) **Face validity** refers to the extent to which a measure appears on the surface to measure what it is suppose to measure. Face validity (sometimes called surface validity) is probably the most commonly discussed type of validity.

(b) **Criterion validity** is a way of assessing validity by comparing the results with another measure. For example, we could compare the results of an IQ test with school results. If the other measure is roughly compared at the same time we call this concurrent validity. If the other measure is compared at a much later time we call this predictive validity.

(c) **Construct validity** is a way of assessing validity by investigating if the measure really is measuring the theoretical construct it is suppose to be. For example, many theories of intelligence see intelligence as comprising a number of different skills and therefore to have construct validity an IQ test would have to test these different skills.

2. There are two main ways of assessing the validity of a procedure:

(a) **Internal validity** is related to what actually happens in a study. In terms of an experiment it refers to whether the independent variable really has had an effect on the dependent variable or whether the dependent variable was caused by some other confounding variable.

(b) **External validity** refers to whether the findings of a study really can be generalised beyond the present study. We can break external validity down into two types.

**Population validity** - which refers to the extent to which the findings can be generalised to other populations of people.

**Ecological validity** - which refers to the extent to which the findings can be generalised beyond the present situation.

\*\*\*\*\* End of extract \*\*\*\*\*

## 9.6 Resume

After looking in depth at validity it is difficult to realise that it is just one aspect of the research project. The Diagram on the next page from a book concerned with monitoring education (Fitz-Gibbon, 1996) considers the issues that need to be considered when developing any type of measurement. She feels there are four main areas to consider:

- Reliable - accurate and consistent
- Valid - measures what it is supposed to.
- Economic - time, effort and cost appropriate
- Reactive- Impact on students, staff, subjects and society

The diagram on the next page considers each of these aspects in more detail.

### Exercise:

Annotate and extend the diagram to reflect what you know about validity now. You may feel that the items under the Economic and Reactive headings do not apply to you?

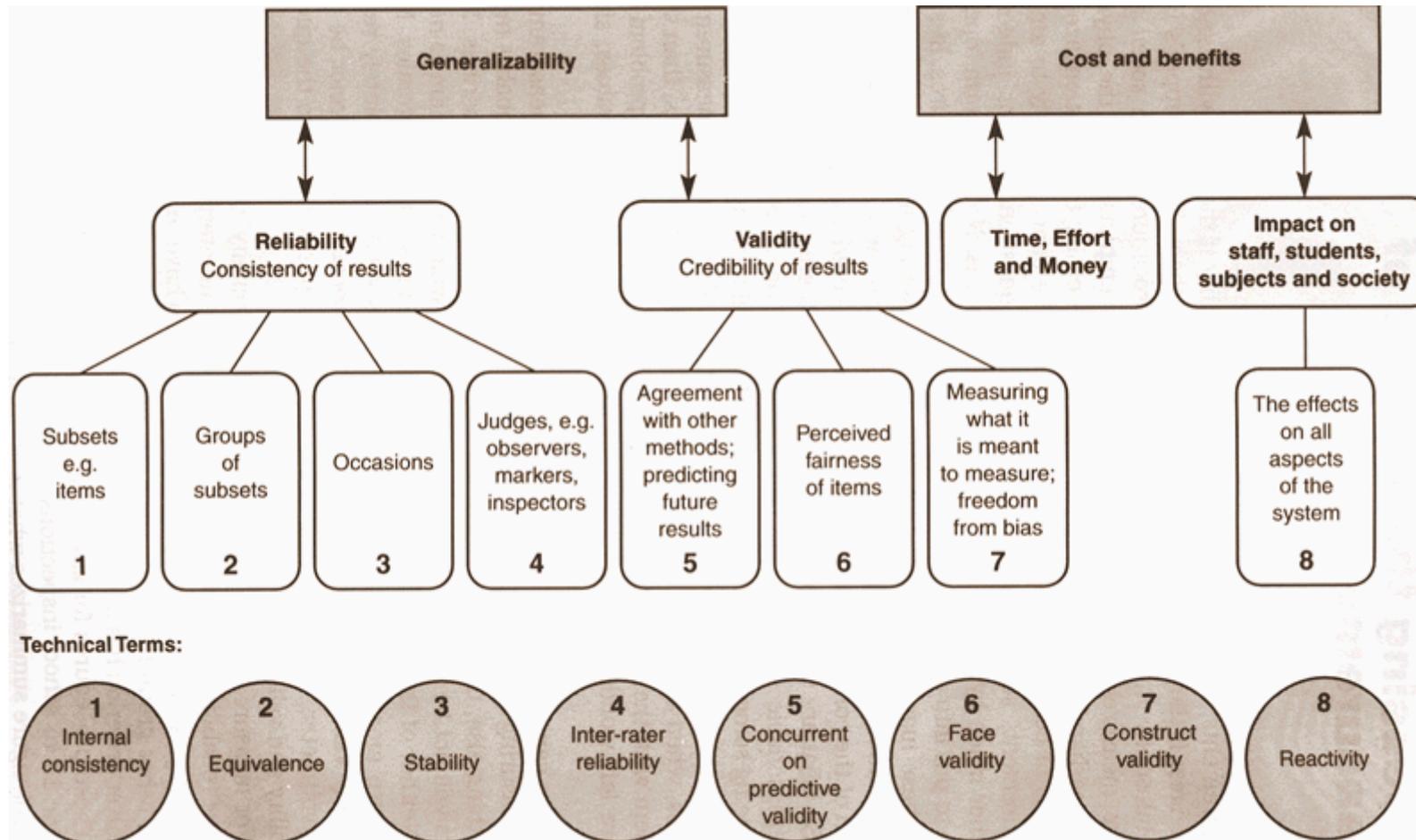


Figure 13.1 Criteria for evaluating the quality of measurements

# 10. Experimental Control

Information for this chapter has been taken partly from Portney & Watkins chapter 9 experimental control.

As mentioned earlier when the researcher measures the dependent variable besides measuring the direct effects of the independent variable she is also measuring to a certain extent '**noise**'. This noise is the result of **nuisance variables that is** those that have not been controlled. These nuisance variables are also called **extraneous variables**. If one, or more, of then happens to effect the dependent variable it is also called a **confounder**. An example was given in the section on one shot correlational studies.

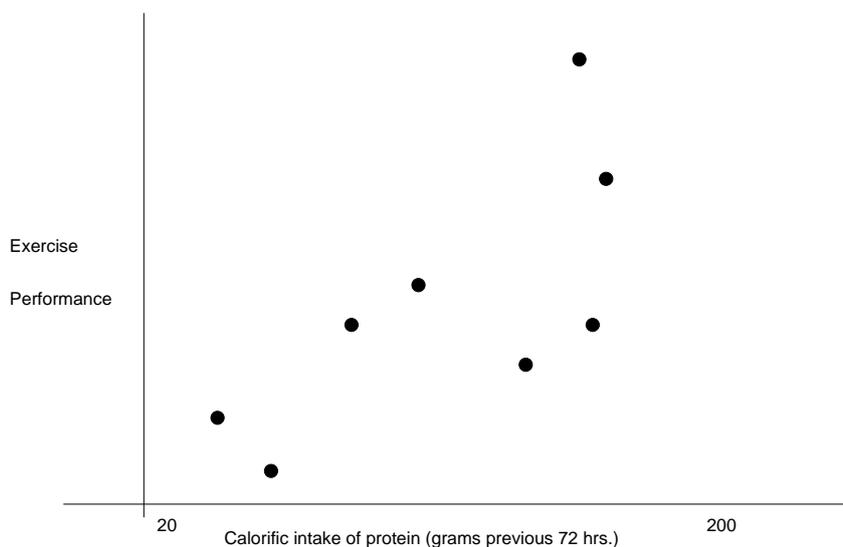
In an experimental design the aim of the researcher is to either remove or control extraneous variables. Control means to ensure they will affect all groups equally. **Bias** means that there is a unintended difference between groups.

Experimental control is achieved by:

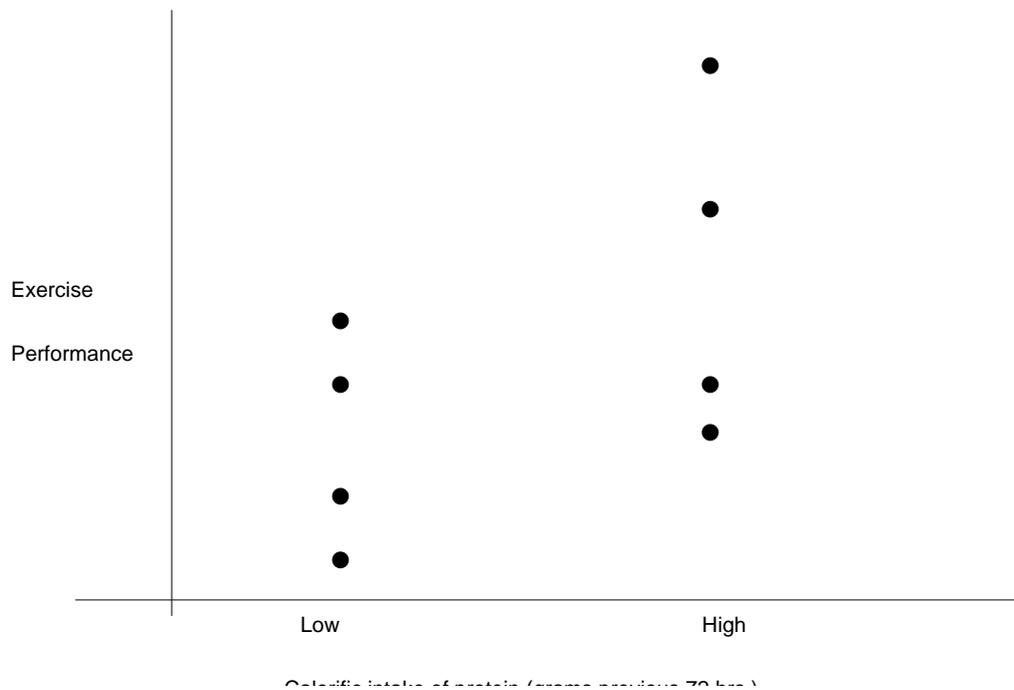
- Manipulation of variables
- Random Assignment [Allocation]
- Control / Placebo Groups
- Protocols
- Blinding
- Controlling / Measuring Inter-subject differences:
  - Homogenous Subjects within groups
  - Blocking
  - Matching
  - Analysis of covariance

## 10.1 Manipulation of Variables

This is when the researcher deliberately manipulates the independent variable. The independent variable can be of any measurement scale but is often ratio or nominal. A ratio example might be the manipulation of daily calorific intake in the form of protein upon exercise performance (the dependent variable). The subject would vary their protein intake as instructed by the researcher and then their exercise performance would be measured. The results might be presented in the following format:



The effect protein intake had upon exercise performance would also be assessed by considering protein intake to be a nominal independent variable, in this instance the nominal variable could have two levels 'low' and 'high'. A group of subjects with comparable exercise scores could be placed in low and high protein diets by the researcher and then their performance measured after a suitable period. This time the results might be presented in the following format:



It is important to note that it is the researcher that has control over the independent variable. She manipulates it then observes a change in the dependent variable. This is very different from the researcher passively measuring protein intake and exercise performance because in this situation she does not know:

- A. if anything else has also effected the exercise performance
- B. Was it the exercise that make the subject eat protein or was it the other way round? Or does there just happen to be a chance association?

By the researcher directly manipulating the independent variable these two alternative explanations can be discounted.

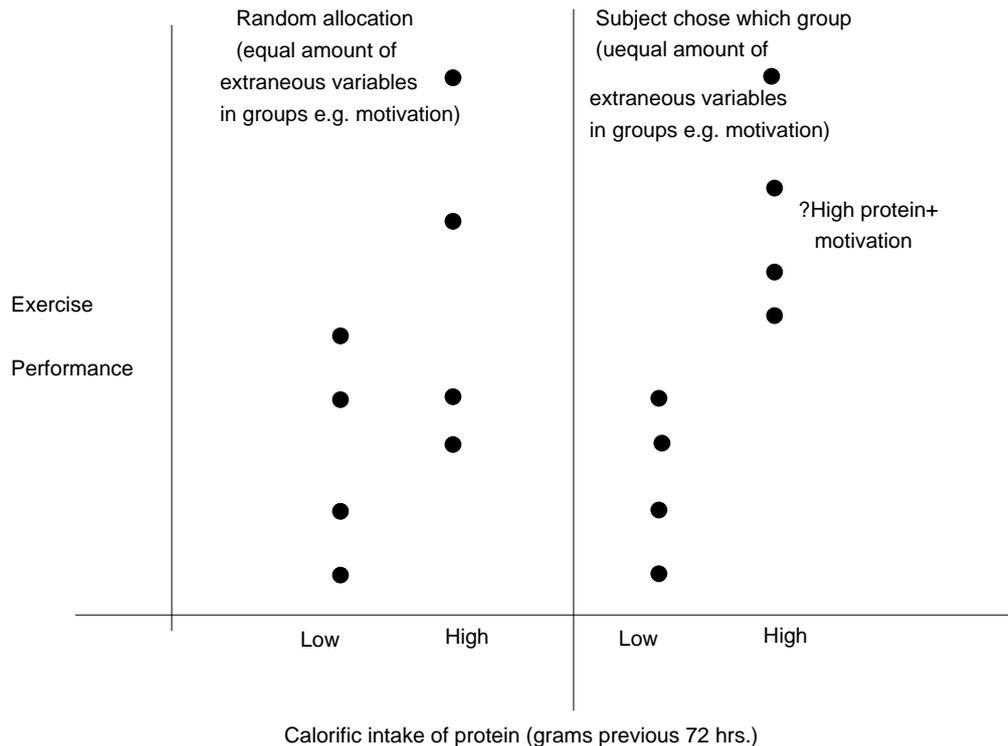
## 10.2 Random Assignment [Allocation]

This is carried out to help ensure group equivalence. Besides ensuring equivalence on the dependent and independent variables it more importantly ensures that each group will have an equivalent spread of confounding/extraneous variables. Continuing with the above example concerning protein intake (as a nominal scale of measurement) and exercise performance. Consider two scenarios:

A. Subjects choose which group to go into. The high protein group will probably contain those subjects which wish to do well e.g. those that are more motivated. Subsequently motivation would probably act as a confounder to exercise performance in the high protein intake group only.

B. Subjects randomly assigned to groups. Now there is equal probability that both groups will contain the same degree of motivation.

The two above situations can be represented graphically:



Random allocation can not be over emphasised. A researcher can always measure everything she believes that may affect the independent variable to check the groups are equivalent, but she can never be sure that she has checked everything. Random allocation, if the groups are sufficiently large, controls for these **unknowables** as well. It is important to have sufficiently large groups if you believe that a particular extraneous factor is relatively rare, to allow it to be adequately represented. Alternative techniques for ensuring you include KNOWN confounders are provided further on.

Consider random assignment like an insurance policy, you do not know if you are going to be burgled but it protects you against it if it should happen.

### 10.3 Control / Placebo Groups

The most effective design strategy for ruling out extraneous effects is the use of a control group against which the experimental group is compared (Portney & Watkins p.128). This may be a little strong.

The process of providing the 'treatment' in itself may affect the dependent variable. To control for this a placebo is often used. For example in the high protein diet group in the above example their performance may have increased because of the additional magnesium that is also found in meat. The low protein group should therefore possibly be given the equivalent dosage of magnesium or alternatively the magnesium removed from the high protein diet. The aim is therefore to make the groups equivalent in all things EXCEPT the treatment variable. Therefore in drugs trials often water or sugar pills are given to the control group to ensure that nothing is different except the drug.

### 10.4 Protocols

The protocol is a document that provides detailed descriptions of how the research is to be carried out.

It will provide details of operational definitions of the independent and dependent variables along with details of strategies undertaken to control for extraneous variables will be given and details of testing schedules along with any instruments used.

Limitations with regards to external validity will be noted. For example noting that the sample consists of over 70 year olds. The main purpose of the document is to ensure standardisation of subject testing and allow replication of the experiment if necessary.

## 10.5 Blinding

There is a great potential for experimenter, and/or subject bias to affect the dependent variable. Researchers may encourage the high protein group to push themselves further when carrying out the performance tests. Similarly subjects in the control group may feel like second class citizens.

Both the above situations can be avoided by blinding. '**First level**' blinding is when usually the researcher carrying out the testing is blinded but not the subject. '**Double blinding**' is when in addition the subject is unaware if they are receiving the treatment or not, in this instance it is usually necessary to provide a placebo. '**Third level blinding**' is when, additionally, those carrying out the data preparation / analysis are unaware of the groups

## 10.6 Controlling / Measuring Inter-subject differences:

The above techniques control for knowable and unknowable extraneous variables. The following techniques control only for knowable cofounders.

### 10.6.1 Homogenous Subjects

This means choosing subjects who have the same characteristic of the extraneous variable. For example in our protein / exercise performance research we may have reason to believe that women may react differently, on the dependent variable, to men when given a high protein diet. The control for this the researcher could only choose only males or females to take part in the research. She has effectively removed the confounder.

### 10.6.2 Blocking

Rather than removing the possible confounding effect of gender in the experiment described above the researcher may decide to measure it by having a group ('**block**') of males and a group of females and then compare the results.

What has happened is that the researcher has included the extraneous variable in the research design as another independent variable, although she has not strictly controlled it this time. Furthermore, the extraneous variable is divided up into **Blocks** (these are just levels of a nominal measure). This type of independent variable is often called a **blocking variable**.

**Exercise:** How might you control for the possible extraneous effect of age in an experiment?

### 10.6.3 Matching

This is when the subject is either matched with another subject in another group or alternatively acts as their own control by being measured repeatedly (= **repeated measures = Crossover** design). The subjects are paired on a extraneous variable, for example if the researcher believes age will act as a confounder subjects in the control group are paired with subjects of a similar same age in the treatment group(s). Alternatively a subject is first subjected to the treatment then the placebo.

#### 10.6.3.1 Problems with Repeated Measures Designs

There are several dangers specific to repeated measures designs:

**Carry-over effect** - This effect occurs when a new treatment is administered before the effect of the previous treatment has worn off.

**Latent effect** - This is where one treatment may activate the dormant effect of the previous treatment or interact with the previous treatment. If this is suspected it is best to avoid a repeated measures design.

**Learning effect** - This is when the response to the dependent variable improves merely due to repetition. Learning effects can be assessed by including a control/ placebo group that perform the same tests repeatedly but not the treatment.

### 10.6.3.2 Controlling Problems with Repeated Measures Designs

Special designs are available that allow assessment of the carry-over effect Cochran & Cox 1957 Experimental design John Wiley & Sons. Possible carryover effects should be minimised by ensuring adequate breaks between treatments.

**Randomised sequencing (Counterbalancing)** - That is ensuring each subject carries out the various treatments in a randomised order. Because of this the carryover and learning effects are controlled.

### 10.6.4 Analysis of Covariance (ANCOVA)

This is a mathematical method which allows those extraneous variables which have been measured at the start of an experiment to be taken into account when analysing the dependent variable at the end of the experiment. It is a complex procedure and often inappropriately used. Howells 1992 Statistical methods for psychology 3rd ed. p.549 -561 suggests that the procedure is frequently used inappropriately.

# 11. Evaluating Research Reports

**None of the above 'check lists' are complete and some may be even considered incorrect in places. A good exercise would be for you to design your own check list after reading these?**

Numerous guidelines are available. A few are given below.

## 11.1 Robin Beaumont 1995

<b>What is the research about</b>	Focus / concern / problem well defined? Theoretical assumptions implicit or explicit
<b>Links with previous research</b>	Previous research discussed Relationship with previous research (methodology, data collection, analysis, discussion, conclusion)
<b>Why was the method chosen</b>	How does method relate to concern of research How does it relate to previous research
<b>Aims of the research</b>	Descriptive Theory testing (deductive ) Generate a theory (inductive) Discovery of significant variables Identify relationship between variables (correlation) Causality (experimental research) Solving a professional problem
<b>Internal Validity</b>	Consider each of the threats identified by Cook & Campbell 1979 (see chapter on validity)
<b>Statistical Validity</b>	Consider each of the threats identified by Cook & Campbell 1979 (see chapter on validity) .... ...
<b>Validity of Measurements</b>	Operationalisation Consider each of the threats identified by Cook & Campbell 1979 (see chapter on validity)
<b>External Validity</b>	Consider each of the threats identified by Cook & Campbell 1979 (see chapter on validity) .... ...
<b>Reliability</b>	See chapter .....
<b>Usefulness of research</b>	Original / replication Practical significance versus statistical significance Application areas

## 11.2 Phillips L R 1986<sup>9</sup>

Theoretical Perspective	Logical adequacy Empirical adequacy Parsimony (Ockam's razor: entities are not to be multiplied beyond necessity)
Review of relevant literature	relation with previous work strengths in current knowledge gaps in current knowledge
Hypotheses or research questions	focus for the investigation anticipated relationships null hypothesis testable
Research design	exploratory descriptive experimental errors - subject, investigator, instrument, procedure validity reliability
Methods and instrumentation	interviews, observation, questionnaires, scales critical incidents
Sample and setting	Size type (random / convenience) etc.
Statistical methods	descriptive inferential
Results	Relationship to hypothesis tables and figures nature of presentation Generalisations possible (external validity)
Discussion of results and implications	nature of conclusions nature of implications recommendations for further study
Ethical considerations	informed consent respect privacy credit where credit due mental or physical discomfort

## 11.3 Oxman A D 1994

This 'checklist for review articles' (BMJ 1994 309 648-51) was presented at a meeting on systematic reviews organised jointly by the BMJ and the Cochrane Centre. It contains a good set of references.

Problem Formulation	Is the question clearly focused?
Study Identification	Is the search for relevant studies thorough?
Study Selection	Are the inclusion criteria appropriate/
Appraisal of studies	Is the validity of included studies adequately assessed?
Data collection	Is missing information obtained from investigators?
Data synthesis	How sensitive are the results to changes in the way the review is done?
Interpretation of results	Do the conclusions flow from the evidence that is reviewed?  Are recommendations linked to the strength of the evidence?  Are judgments about preferences (values) explicit?  If there is 'noevidence of effect' is caution taken not to interpret this as 'evidence of no effect'?  Are subgroup analyses interpreted cautiously?

<sup>9</sup>Phillips L R F. 1986 A clinicians guide to the critique and utilisation of nursing research. Appleton-Century\_Crofts, Norwalk, Connecticut.

The paper also provides a table relating to 'levels of evidence for treatment':

Level 1	The lower limit of the confidence interval for the effect of treatment from a systematic review of randomised controlled trials exceeded the clinically significant benefit.
Level II	The lower limit of the confidence interval for the effect of treatment from a systematic review of randomised controlled trials fell below the clinically significant benefit (but the point estimate of its effect was at or above the clinically significant benefit).
Level III	Non-randomised concurrent cohort studies
Level IV	Non-randomised historical cohort studies
Level V	Case series
Detailed definitions for these levels of evidence and corresponding grades of recommendations are provided in:	Cook DJ Guyatt GH Laupacis A Sackett DL. 1992 Rules of evidence and clinical recommendations on the use of antithrombotic agents. Antithrombotic Therapy Consensus Conference. Chest 102 305 - 11S

## 11.4 Hawthorn P J, 1983

Unfortunately I have lost the detailed reference for this set of criteria, I think it was from a medical nursing source any help in relocating it most welcome.

Criteria	If undertaking a research project	To Evaluate a research report
<b>Formulating the problem</b>	Is the examination of the problem worthwhile? Why? To Whom?	Is the problem or purpose of the study clearly stated?
<b>Limitations</b>	How much will the project cost? Who will pay? How long will it take? Is the study within my capacity (e.g. skills)?	How was the project financed? (Are the results biased because of the interest of the financing body?) Who undertook the work? Does it seem appropriate?
<b>Steering Committee</b>	Is it possible to form such a committee even if not required by the funding body?	Was there a steering committee? Which experts were consulted?
<b>Literature</b>	What background knowledge is required? What previous attempts have been made to examine this particular problem or similar ones?	Does the researcher appear to know his subject? Is there discussion of related research and are both positive and negative arguments presented?
<b>Theoretical concepts</b>	What framework is needed to place the research into context?	Is it clear whether a theoretical framework has been used?
<b>Assumptions</b>	Have any assumptions been made? What explanation is needed?	Are assumptions made? Is their use explained? Are they justifiable?
<b>Type of study</b>	What type of study is being attempted? (descriptive or experimental?)	Is the study described adequately?
<b>Hypothesis</b>	Hypothesis is necessary if study is of experimental design	Do hypotheses follow logically from the original problem?
<b>Aims/questions</b>	Have aims/questions been clearly stated? Are variables identified? Have findings from the literature been used in order to formulate these aims and questions?	Do the aims and questions which are posed follow logically from the original problem? Can the major variables be identified?
<b>Population</b>	Who or what forms the population for investigation? (see ethical considerations)	Is the chosen population / group appropriate?
<b>Sample</b>	Is the sample representative of the population? Is the size of the sample appropriate in terms of time available and cost? Will the size of the sample be sufficient to give the results some reliability? What is the method of selection? What is the 'response rate'? Will the validity of the findings be established?	Is the method of selecting the sample clearly stated? Are the reasons for this type of selection given? Of whom and how many is the sample composed? Does it seem appropriate? Is the 'response rate' stated?

<b>Methodology of tools</b>	<p>What method of data collection is to be used? Have the reasons for the choice been explained? Has the choice been limited because of the cost or time available?</p> <p>Possible methods: Interviews - unstructured, semi-structured, structured. Questionnaires - open / closed questions. Likert scales multiple choice questions etc. Observation - participation / non-participation, continuous or interrupted. Work study and activity sampling techniques can be used. Critical incidents may be noted. Quantitative and /or qualitative data may be sought. 'Diary keeping' 'Q sorting'. The use of tape recording, audio videotape are further possibilities</p>	<p>Are the reasons for choice of method used given? Are the methods - advantages &amp; disadvantages - discussed? Are the copies of the interview schedules and questionnaires provided with the report?</p>
<b>Ethical Considerations</b>	<p>Is the proposed method ethically acceptable? Who needs to be consulted (e.g. the local ethical committee?) Has permission to carry out the study been obtained from the necessary people? Will the participants involved be given an explanation of what is to happen and the option to be involved or not?</p>	<p>Has the researcher considered the ethics of the research proposal? Was the necessary permission to undertake the project obtained? Did the participants have the opportunity to refuse to participate?</p>
<b>Pilot Study</b>	<p>It is necessary to check errors of methodology. Is the sample correct? Is the information received relevant to the questions to be answered? Does the method of analysis appear to be satisfactory? What modifications are necessary?</p>	<p>Has a pilot study been completed? What modifications were made and why?</p>
<b>Analysis</b>	<p>How are the facts or opinions which have been collected to be utilised? Can they be categorised by hand or are more sophisticated methods required? Cope-chat cards, automatic sorters or computer analysis may be considered. What statistical tests are appropriate?</p>	<p>Is the method of analysis understandable to the reader? Are reasons for choice of statistical test given? Is the statistical probability of results by chance given?</p>
<b>Results</b>	<p>These must be kept separate from the conclusions. An exact account of the information obtained should be given. Present the results in an appropriate way for the readers of the report. Consider the use of tables, histograms and graphs.</p>	<p>Are these intelligible to the reader and relevant to the problem? Are 'raw' figures always given if results are expressed in percentages? Are graph scales explained? Are statistical results and the probability of significance included? Are the tables helpful?</p>
<b>Conclusions</b>	<p>Important part of the report Must be based on obtained results, even if findings contract expectations. Relate the findings to questions posed originally and to the existing body of knowledge and relevant theory.</p>	<p>Do the conclusions relate logically to the results? Are the aims and the questions posed earlier answered? Has the hypothesis been proved or rejected? What omissions have been made? Has the researcher referred to these?</p>
<b>Recommendations</b>	<p>What changes might arise as a result of these inquiries? Which questions remain unanswered? Which need to be reconsidered in the light of the results? Is further research suggested?</p>	<p>Are the recommendations self-evident after reading the rest of the report? Should you, as a reader, be attempting to implement them?</p>
<b>Timetable</b>	<p>Allow approximately one third of the time for reading, one third for field work and one third for writing up. Writting up: check style requirements for the report. Probably three or four drafts will be necessary</p>	

## 11.5 Greenhalgh 1997

Trisha Greenhalgh wrote a series of articles about reviewing research papers the references are given below.

### [Trisha Greenhalgh](#)

How to read a paper : getting your bearings (deciding what the paper is about)  
BMJ, Jul 1997; 315: 243 - 246.

### [Trisha Greenhalgh](#)

How to read a paper: Assessing the methodological quality of published papers  
BMJ, Aug 1997; 315: 305 - 308.

### [Petra M Boynton and Trisha Greenhalgh](#)

Selecting, designing, and developing your questionnaire  
BMJ, May 2004; 328: 1312 - 1315.

### [Trisha Greenhalgh and Rod Taylor](#)

How to read a paper: Papers that go beyond numbers (qualitative research)  
BMJ, Sep 1997; 315: 740 - 743.

### [Cynthia K Russell and David M Gregory](#)

Evaluation of qualitative research studies  
Evid Based Nurs 2003 6: 36-40.

### [Trisha Greenhalgh](#)

How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests  
BMJ, Aug 1997; 315: 364 - 366.

### [Trisha Greenhalgh](#)

How to read a paper: Statistics for the non-statistician. II: "Significant" relations and their pitfalls  
BMJ, Aug 1997; 315: 422 - 425.

### [Trisha Greenhalgh](#)

How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses)  
BMJ, Sep 1997; 315: 672 - 675.

### [Trisha Greenhalgh](#)

How to read a paper: Papers that tell you what things cost (economic analyses)  
BMJ, Sep 1997; 315: 596 - 599.

### [Trisha Greenhalgh](#)

How to read a paper: Papers that report diagnostic or screening tests  
BMJ, Aug 1997; 315: 540 - 543.

### [Trisha Greenhalgh](#)

How to read a paper: Papers that report drug trials  
BMJ, Aug 1997; 315: 480 - 483.

## 12. Questionnaire Design an Introduction

For this section, please see section 14 on the main web site:

<http://www.robin-beaumont.co.uk/virtualclassroom/contents.htm>

## 13. References

BMJ editorial 1994 checklist for review articles 309 648-51

Campbell, D T 1957 Factor relevant to validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.

Campbell, D T, Stanley, J C 1966 *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cook T D, Campbell, DT 1979 *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Dixon W J Massey F J. 1983 *Introduction to Statistical Analysis* p.209 4th ed. international student edition. McGraw-Hill Book Co. London.

Fitz-Gibbon Carol (1996) *Monitoring education; Indicators, Quality and effectiveness*. Cassell [An excellent book if you want to take research, not just of the educational type, seriously. It manages to describe various types of research as well as validity and complex statistical techniques with unnerving simplicity and clarity. A joy to read.]

Giddens A 1989 *Sociology* p.663 Cambridge Policy Press in association with Basil Blackwell

Graziano A M, Raulin, M L 1993 *Research Methods: A Process of Inquiry*. New York: HarperCollins College Publishers.

Hart S Staveland L 1988 *Development*

Hilliard D C. 1984 Media images of male & female professional athletes: An interpretive analysis of magazine articles. *Sociology of Sports jour.* 1 251 – 262

King S 1993 The politics of the body and the body politic: Magic Johnson and the ideology of AIDS. *Sociology of Sports jour.* 10 270 – 285

Lecompte M D Goetz J. P. 1982 Problems of reliability and validity in ethnographic research. *Review of Educational Research*, 52, 31-60.

Marrow J R. Jackson A W. 1993? How significant is your reliability. 64 3 352-355. [Provides tables for the significance of a reliability measure].

Masters KS Ogles MB Jolton JA 1993 The development of an instrument to measure motivation for marathon running. *Research Quarterly for exercise and sport* 64 (2) 134 – 143

Nixon H.L 1991? Sports sociology that matters: Imperatives and Challenges for the 1990s *Sociology of sport jour.* 8 281-294 and Harris J C 1989? Suited up and stripped down: perspectives for sociocultural sport studies *Sociology of sport jour.* 6 335-347

Oppenheim A N 1992 *Questionnaire design, interviewing and attitude measurement*. London, UK: Pinter Publishers.

Phillips L R F. 1986 *A clinicians guide to the critique and utilisation of nursing research*. Appleton-Century\_Crofts, Norwalk, Connecticut.

Portney & Watkins 1994 *Foundations of clinical research*. Appleton & Lange

Portney L G Watkins M P 2000 Foundations of Clinical Research: Applications to Practice. Ed.2 Upper Saddle River, NJ, Prentice-Hall.

Redican B. Hedley D S. 1988 A field studies project in a City Health & Leisure club. Sociology of Sports jour. 5 50 - 62

Rintala J Birrell S. 1984 Fair treatment for the active female: A Content Analysis of 'Young Athlete' (YA) Magazine. Sociology of Sports jour. 1 213 – 230

Robson 1993 Portney & Watkins 1994 Foundations of Clinical Research: Applications to Practice. Ed.2 Upper Saddle River, NJ, Prentice-Hall.

Robson, C 1993 Real world research: A resource for social scientists and practitioner researchers. Oxford: Blackwell.

Smith H W. 1975 Strategies of Social Research. p101 London Prentice Hall International.

Stanton A Glantz 3rd ed. 1992 'What does not significant really mean' in Primer of Biostatistics by McGraw-Hill.

Sugden, John 1993 The exploitation of disadvantage: the occupational sub-culture of the boxer' in sport, leisure and social relations John Horne, David Jary, Alan eds. Tomlinson pub. Routledge & Kegan Paul p.187 - 209

Taylor J 1987 A review of Validity Issues in Sport Psychological Research. Journal of Sport Behaviour 10, 1 3-12

Waller J 2002 Fabulous science: Fact and fiction in the history of scientific discovery. Oxford University Press (pp. 78 – 99).

Zoltie N Dombal FT de 1993 The hit and miss of Injury severity score (ISS) and combined Trauma and Injury Severity Score (TRISS). BMJ 307 906 - 9 (October 9th)

# 14. Index

- Case Study, 5
- Classic Experimental Design, 9
- Client, 4
- Cohort, 4, 11
- confounding of the dependent variable**, 13
- Consistency(see reliability), 15
- Constructs, 16
- Control Group, 4
- Correlation, 8
  
- Data, 4
- Data Collection Schedules - things to consider, 10
- dependent variable, 17
- dimensions**, 16
- double blind experimental design, 4
  
- Ethnographic Studies, 5
- Evaluating Research Reports**, 40
- Ex Post Facto, 13
- Experimental bias, 31
- Experimental Control**, 35
  - ANCOVA, 39
  - blinding**, 38
  - controlling / measuring intersubject differences**, 38
  - controls / placebo**, 37
  - Counterbalancing, 39
  - manipulation of variables**, 35
  - randomisation**, 36
  - repeated measures - problems**, 38
- Experimenter, 4
- Experiments, 8
- Experiments, disadvantages of, 11
- Exposure Randomisation, 11
  
- Factorial Design, 10
- Field Methods, 11
  
- Group, 4
- Groups [units], 10
  
- Hawthorn effect, 31
- Hypotheses**, 22
- Hypothesis**
  - alternative**, 27
  - null - why**, 25
  - Null (h<sub>0</sub>)**, 24
  - relationship with inferential statistic**, 26
  - research / scientific**, 22
  - statistical hypothesis**, 23
  - testable**, 24
- Hypothesis Testing, 5
  
- independent variable, 17
- indexes, 18
- Inductive Approach, 6
- Instrument, 4
- Instrumentation, 10
- Interrupted Time-Series Design**, 11
  
- Intervention, 4
- Investigator, 4
  
- lag/response time, 17
- Learning effect, 38
- Longitudinal Studies, 11
  
- Manipulated variable, 4
- measurement
  - assesment of, 33
- multidimensional**, 16
- Multiple Interventions, 10
- Multiple Time-Series Designs**, 11
- Multiple treatment interactions, 31
  
- Naturalism, 5, 8
- Non Test Group, 4
- Nonequivalent Control Group Design**, 12
- Nonexperimental Designs, 13
- Non-Experiments, 8
- Nonrandomised Groups, 12
  
- One-Group Pretest-Posttest Design, 14
- One-Shot [Cross-sectional] Study [Case Study][Survey]**, 13
- One-Shot Correlational Study**, 13
- OPCS, 6, 11
- Operationalisation, 10, 16, 21
  - GRAPHICAL graphical method of representing**, 19
  
- Placebo, 4, 9
  - experimental control**, 37
- Protocols**, 37
  
- Qualitative Research Methodology**, 5
- Quantitative Methodology**, 5
- Quasi Experiments, 8
  
- Randomisation**, 8
- Reliability**, 15
  - Cronbachs Test, 15
  - Equivalent [alternative/multiple/pure] forms [measures], 15
  - Influences on, 15
  - Inter-rater, 15
  - Relationship with validity, 15
  - Split half random allocation, 15
  - Testing**, 15
  - Test-Retest, 15
- Replication - of Groups, 10
- Researcher, 4
  
- Sample, 4
  - randomisation / assignment / allocation**, 36
- Sample - Things to Consider, 10
- Scales**, 18
- Secondary data, 4
- Separate-Sample Pretest-Posttest Control Group Design, 12

**Separate-Sample Pretest-Posttest Design, 12**

single blind experimental design, 4

Solomon Four-Group Design, 9

**Static-Group Comparison Design, 14**

Structured Analysis, 5

Subject, 4

Subject Reactivity to Pretesting, 9

target population, 23

Test Group, 4

Tolerance levels, 15

Treatment, 4

validity, 15, 20, 28

concurrent, 32

consequential, 32

construct, 20, 32

**construct - threats to, 31**

content, 20

convergent, 32

criterion, 32

curricular, 32

discriminant, 32

Ecological, 33

**external - generalisations, 32**

face, 32

face (logical), 20

internal validity - threats, 28

**of findings, 28**

**of statistics (statistical validity), 30**

population (type), 33

predictive, 32

**Relationship with Reliability, 20**

**types of, 20, 28**

validity of findings

external validity, 12

history, 12

Internal, 12

maturation, 12

variable

confounder, 35

variables

Inadequate definitions, 31

nuisance, 35

Volunteer, 4

Wittgentein, 6

**Document info:**

Robin Beaumont

Source: Laptop; C:\Hlcourseweb new\chap16\SEMBK2.DOC Time: 13/04/00