YOUR PSYCHOLOGY PROJECT HANDBOOK

BECOMING A RESEARCHER

Second Edition

Clare Wood, David Giles & Carol Percy

ALWAYS LEARNING

PEARSON

YOUR PSYCHOLOGY PROJECT HANDBOOK

Visit the *Your Psychology Project Handbook, second edition* Companion Website at **www.pearsoned.co.uk/wood-psychproject** to find valuable **student** learning material including:

- Downloadable forms, checklists, templates and worksheets to help you in the different stages of your project work
- Annotated examples of good and bad reports
- Guidance on how to organise a literature review
- Self-test questions to check your understanding
- Annotated links to relevant sites on the web

However, if we are just interested in the main effects, we may as well simply run two *t*-tests. The whole point of a factorial experiment is that you have set it up to explore an interaction – in other words, when the effect of one factor on another factor varies according to the different *levels* of that factor. In a simple 2 × 2 design such as this, each factor only has two levels, so interpretation is fairly straightforward. Adding sound to still pictures (level 1 of the movement factor) has little effect. Adding sound to moving pictures (level 2), however, boosts recognition by what could well be a significant amount.

In the ANOVA calculation, this contrast between the effects of sound at the two levels of movement would probably constitute *a significant interaction* between the factors of movement and sound. While interpretation is easy with a simple design, for more complex designs (where you have more than two factor levels), you would need to carry out post-hoc tests, such as Scheffé, or Bonferroni corrected *t*-tests at each of the factor levels, in order to fully interpret the significant interaction.

MAXIMISING MARKS

Keeping it simple

Students often get carried away at the design stage, wanting to add a lot of independent variables in the mistaken belief that complex designs automatically ensure higher grades. In reality, this is seldom the case, usually because not enough thought has been put into the analysis of that design, which means that the student is unable either fully to analyse or interpret the data. A good piece of advice when it comes to factorial designs is to minimise the number of within-subject factors (independent variables) in your design and, where you do have them, try to keep the number of levels in those factors to a minimum (ideally just two conditions). This is because, if you have too many levels and/or too many within-subject factors, post-hoc analyses become very difficult indeed. Elegant designs, well-justified and explained, are more likely to attract higher marks than overly complex and poorly explained designs. Think about the studies that you consider to be 'classics': usually these have very basic designs that are simple and test a key idea (not ten!).

4.2.3 Balancing and checking

There are a number of important checks that you will need to carry out when designing an experiment. First of all, if you are presenting your participants with more than one single task or measure, you will need to consider some form of *counterbalancing* the order of presentation in order to eliminate potential order or practice effects.

Let us return to the performance anxiety example from Section 4.2.1. You are interested in the effect of audience size on task performance and so you need to collect a number of measures of anxiety. You could use a self-report measure but you also want some physiological data, so you fit your participants with a BioPac and measure heart rate and galvanic skin response.

You also want to assess performers on a variety of tasks. Suppose you have three different manual tasks: one involving balance; another involving hand/eye coordination; and a target-hitting exercise. It is possible that these tasks differ with respect to difficulty. It is also possible that a good performance on one task facilitates performance on the next (through elevated self-confidence or whatever). Therefore, the order participants carry them out may accidentally interfere with the main variable (audience size).

In order to minimise any of these potential confounding factors, it is good practice to randomise the order that participants carry out the tasks, so that task difficulty, practice or any other cumulative

advantage cannot be held accountable for your eventual results. One way of doing this is to use a *latin square*. A latin square is an array of symbols in which each symbol occurs once in each row and once in each column. You start by drawing up a list of your participants, and then work out the total set of possible combinations of tasks or measures. Table 4.2 shows a latin square for the first three participants in our performance anxiety experiment.

TABLE 4.2 Latin square for performance anxiety task.

Participant	Order of task		
1	Balance	Coordination	Target
2	Coordination	Target	Balance
3	Target	Balance	Coordination

This latin square has not exhausted all possibilities of order. How about target-coordination-balance? In fact there are 12 different possible latin squares that you could have come up with for this particular (3 × 3) array. You could create a new latin square for each set of three participants and wait until participant number 37 before starting all over again. Or you could, if you prefer, simply repeat this latin square *ad nauseam*, so that participant 4 follows participant 1's order, and 5 follows 2, and so on.

Another important experimental check is known as a *manipulation check*. This is a way of finding out if your manipulation has been successful; in other words, it has produced the effect that you were after. In studies where you want to manipulate mood it is essential. Imagine you wanted to induce sadness by playing a sentimental piece of music. You could not simply assume that the participants in the 'sad' condition would be feeling sad because the piece of music you have chosen to play them makes *you* feel sad! Here, the manipulation check might be nothing more than a self-report mood inventory that participants fill in after the task or measure is completed. Ideally this is the sort of thing that you would check during the pilot study – it will be too late to change the music, if you leave it till the main data collection phase!

4.2.4 How many participants?

In recent years, psychologists and, perhaps more importantly, sources of research funding have become increasingly concerned about the **statistical power** of their studies. Jacob Cohen and other figures (see Cohen, 1994) have claimed that many (typically experimental) psychology studies are 'underpowered' – in other words, they do not involve enough participants to be confident of detecting significant differences. (Incidentally, low power does not mean that a study will always be non-significant, but it means that you might overlook a real difference.) There is now some very good software available that you can download (for free!) in order to carry out an *a priori* power analysis. This means that the program will calculate how many participants you need in each group in your study for it to be sufficiently powerful to detect a significant result. It is called G*Power (version 3.0) and can be downloaded from the University of Dusseldorf's website (Heinrich Heine Universität Düsseldorf, 2008). All you need to specify are the following criteria:

- type of statistical test you hope to run (ANOVA, regression, etc.);
- number of different groups or variables;
- alpha (typically $p < .05$);
- anticipated effect size (unless you are very confident, always choose medium, which is around $d = .5$).

You will usually find that G*Power is somewhat harsh on experimental sample sizes, and rather generous on correlational designs. That has got more to do with the nature of the statistics involved than any deep-seated bias. However, although advocates of power calculation are quite vociferous, one thing that needs to be borne in mind is that, for some studies, you have to accept low statistical power – when studying hard-to-recruit populations, for instance, such as unusual clinical groups. After all, just because a study of 20 very interesting people is low powered, that does not mean it is not worth carrying out!

4.2.5 Hypothesis testing

Of course, one of the major considerations in research design is that your study will enable you to answer your research question. When doing experimental research, this should be a straightforward business if you have designed your study carefully, because you will ask very precise questions indeed: hypotheses that relate your independent variables (or factors) and dependent variables to one another.

Whatever your research question, you will find it helpful if your hypotheses are as explicit as you can make them (so long as you have provided sufficient evidence to support those predictions in your literature review). Unless you are doing genuinely exploratory research, and taking a leap into the unknown, it is best to avoid *two-tailed hypotheses*, as it is very unlikely that someone would go to the trouble of designing an experiment of which they genuinely have *no idea* what the outcome is likely to be.

It is not enough even to say 'there will be an interaction between X and Y'. You should really have set up your experiment to test a specific effect, and be able to articulate what that effect is likely to be. Ultimately, two-tailed hypotheses are for speculators – people who are so determined to obtain a significant result that they refuse to make any clear prediction.

One potential hazard you may have already come across, during research methods or statistics classes, is *Type 1 error inflation*. Type 1 errors are referred to as 'false positives' – in other words, you claim evidence of an effect when there actually is not one. These kinds of errors can be made when reading significance values from large tables of correlations (for example) and pouncing on any spurious association because it has an asterisk next to it. They are always a possibility when you are making multiple comparisons – for example an independent variable with five levels.

There are two methods for dealing with Type 1 error inflation. The first is the *Bonferroni correction* that you apply when carrying out post-hoc significance tests between pairs of means. In a one-way ANOVA with five conditions, your chance of finding a significant pair of means is several times greater than in a *t*-test; not because your design is more powerful, but simply through the laws of probability. Bonferroni's correction states that you divide your alpha (usually .05) by the number of comparisons you want to make in order to obtain a new, more stringent, alpha. In other words, if you want to compare three pairs of variables, you would convert alpha to .0167 (.05 divided by three), so that you would have to consider a p value of .02 as non-significant.

This correction might be seen as unnecessarily harsh in some eyes. Indeed it *is* harsh, because there is a second, preferable way of dealing with multiple comparisons: the *planned comparison*, where you set up your design with the comparisons already predicted in advance, so you do not have to make all those extra comparisons that incur the wrath of Mr Bonferroni. You do this by setting up your SPSS program so that you weight each of the variables according to your hypothesis. So, if you predict that of your three variables one will be highest, one in the middle and the other lowest, you accord those variables the values +1, 0 and −1. SPSS then calculates a formula based on the expectation of this emergent pattern. It is much more powerful, and much more sensitive to actual means differences, than Bonferroni's correction.

4.2.6 Single-case experiments

These are far less common than group-based experiments, but occasionally undergraduates get the opportunity to have access to clinical populations, particularly in an area such as neuropsychology. A single-case experiment is much harder to design than a group experiment because of the restrictions on control – you certainly cannot have a control *group* if you only have one participant!

The classic single-case design incorporates two experimental phases – a *baseline* phase and a *treatment* phase. This is referred to as an *A-B design*. At its simplest, you would take a series of measurements during the baseline phase to establish the pre-treatment level of functioning, and then a longer series of measurements to observe the effect of the treatment.

You can mix up baseline and treatment phases in order to observe the effects of withdrawal (bear in mind that this type of design has serious ethical implications), thereby creating A-B-A-B designs. You could also introduce a second treatment (A-B-A-C). Any combination is possible. Alternatively, you might wish to create a more complicated design in which you observe the effects of treatment on several different behaviours, known as a *multiple baseline* design.

Of course, single-case designs are not restricted to neuropsychological cases. You could construct a similar design to observe the effectiveness of a teaching intervention or relaxation technique. However, these types of study are reliant on you delivering the intervention, so be careful if it is something that requires a degree of training. Your supervisor should be able to recommend the best practice in this instance.

MAXIMISING MARKS

Reporting single-case experiments

To get a sense of how single experiment designs are written up and the data presented, have a look at single participant or small N studies in the *Journal of Applied Behavioral Analysis.* This area of psychology very commonly uses A-B type designs with small numbers of human participants, and you can get a sense of how to present such data effectively.

4.3 CORRELATIONAL DESIGN

In this section, we consider the types of study that rely on correlational data for their analysis – not just correlations but also regressions and other complex designs where you are essentially asking about associations between variables, or the ability of one or more variables to predict performance on another variable.

4.3.1 Basic types of correlational study

In psychology, the vast majority of correlational designs are put together with some form of multiple regression in mind. That is, you begin with a variable of interest (say, physical well-being), and then identify a set of variables that you believe has a collective impact on that variable (typically, a set of psychological constructs or possibly social phenomena). The former is referred to as a *criterion* variable; the latter as a set of *predictor* variables.

In theory, you can have as many criterion variables and predictor variables as you like, but there are some restraints imposed in relation to statistical power: for each predictor you add to a multiple

regression equation, you need to find more participants in order to achieve satisfactory power (20 participants per predictor variable is a good rule of thumb). Also, it is potentially an analytical nightmare to have more than a couple of criterion variables, since you will need to find some way of presenting a whole series of regression equations. For undergraduate project work, it is probably best just to identify *one* criterion and stick with it.

COMMON CONFUSION

Terminology in regression designs

If you are conducting a study that is aiming to predict scores on a measure, you will have a predictor variable (or several) and a criterion variable (i.e. the variable that you are trying to predict). However, some statistical packages will ask you to identify your 'independent' variable and your 'dependent' variable when performing a regression. When this happens, you can think of your predictors as independent variables and your criterion variable as your dependent variable.

There are two basic formats for correlational research. The first is the *cross-sectional* study, which is sometimes referred to as a 'snapshot' approach because it captures your sample at a particular moment in time. The second is a *longitudinal* study, or sometimes a *panel* design, where the same sample is contacted at a later date and then re-tested. *Time* then becomes a key predictor variable in your study.

MAXIMISING MARKS

Pilot studies

Do I need to carry out a pilot study? The answer to this question very much depends on how confident you are in your study design and materials. The rule of thumb is that, the more original the study the greater the need to pilot. If you are doing a study in which all the materials have been used before in published research (e.g. a set of psychometric instruments), then a pilot is not really that necessary. If, on the other hand, you are creating a unique set of stimuli or a new scale, or are interviewing someone on a specific topic or observing a particular behaviour for which there are no standard checklists, you will need to incorporate a pilot phase into the first part of your study. Pilot studies can be used simply to fine tune: to see if questions make sense in an interview schedule, to see if participants can see the visual stimuli on the screen, etc. Or they can be more exploratory: to see if a particular topic generates stimulating interview data, to see if participants can actually perform the task that you want to use to measure your dependent variable (DV). In this latter case, you need to be prepared to scrap the whole idea and start again (as far as the materials are concerned, that is). On the other hand, you may find that the pilot goes swimmingly (if that is not too much of a mixed metaphor!) and that you can use all your pilot data in the final analysis. That is the risk you need to take. An original study without a pilot phase, appropriately reported (see Chapter 10), will probably lose marks. Above all, never say, in your Discussion, something such as 'the failure of the experiment to find significant differences resulted from participant boredom due to the length of the trials', if a good pilot study could have eradicated this potential confound.

Of course, whether you choose a cross-sectional or panel design depends on your research question – if you are examining changes over time (perhaps in response to some kind of intervention, such as an educational initiative or a treatment of some type) then a pretest-post-test design will be your absolute minimum, and you will need at least one follow up study at some later time point. However, few final year undergraduates have sufficient time to prepare such long-term studies and so, for practical reasons above all else, the cross-sectional study is by far the more common at this level.

4.3.2 Other multivariate designs

There is a vast array of multivariate statistical procedures now available to psychology researchers, and it may be that your supervisor has enough confidence in your ability to allow you to design a study that can be analysed by advanced techniques such as structural equation modelling or multidimensional scaling. Such designs are beyond the scope of this chapter, so we recommend that you seek out a specialist text such as Tabachnick and Fidell (2007).

However, you may wish to use correlational analyses to deal with the data collected by a humble questionnaire, where, rather than identifying discrete measures such as self-esteem or anxiety that have their own psychometric scales, you have simply asked a series of isolated questions on a particular topic. Typically, such questionnaires are used alongside more established psychometric measures, in order to collect demographic data, or to explore a particular behaviour that you would like to relate to psychological variables, such as alcohol consumption or media use.

In these situations the challenge is to create meaningful measures out of your individual items. Tips on how to do this well can be found in Chapter 8 (Section 8.4.2).

4.4 OBSERVATIONS

Observational studies fall neatly into two types: **non-participant** (or **remote**) and **participant**. We will begin by discussing the advantages and disadvantages of each in turn.

A *remote observation* is one where the researcher is physically removed from the location where the observed behaviour is taking place. Some research questions demand remote techniques, such as the observation of animals or infant humans, where the presence of the researcher cannot be explained to the observees, and where interference in the behaviour of the group would contaminate the study findings. Another reason for remote observation is the need for large numbers – for instance the study of shopping behaviour in a busy mall might be more efficiently undertaken without any direct contact with shoppers themselves.

Naturally, the limitations of such a study are obvious: you run the risk of collecting superficial data that tell you little about the private cognitions, cultural influences or social motivations that are driving the behaviour. Nevertheless, you are left with the feeling of having captured a 'slice of real life' as it happens in a natural environment.

Participant observation differs in that the observer actually becomes part of the natural environment, in either an *open* or *covert* fashion. Participation is said to be open when the observees are fully aware of the observer's status as a researcher; in a covert study, the observer takes care to conceal their identity, often by 'infiltrating' a group in order to gain an insider's perspective while not apparently interfering with the group's behaviour.

The disadvantages of participant observation appear obvious, particularly in the open type of design, while covert designs contain a multitude of ethical problems for both observers and observees. In practice, participant observations are not nearly as common in psychology as in other social sciences, particularly at postgraduate and postdoctoral levels, where they can be carried out