Oxford Resources for IB
Diploma Programme

IB PREPARED

2025 EDITION

PSYCHOLOGY

Alexey Popov

OXFORD

Contents

3 Research methods in psychology

Assessment tip

The skills related to analysing research methodology and interpreting data obtained in a psychological study will be crucial in all examination components.

- Paper 1 (Section A and B): assessed indirectly through your ability to understand how examples support a concept or a theory.
- Paper 1 (Section C): assessed indirectly through your ability to analyse critically and evaluate psychological ideas based on their empirical support.
- Paper 2 (Section A): directly assessed through a series of questions about research methodology, rooted in the context of your class practicals.
- Paper 2 (Section B): directly assessed because discussing a research study through the lens of a concept almost always requires discussing the link between a study's evidence and the theoretical conclusions derived from it.
- Paper 3 (HL): assessed indirectly through your ability to understand how a study's findings should be interpreted in the context of the research method used.
- Internal Assessment (IA): assessed directly because your research proposal reflects your understanding of how research methods work and which methods are appropriate, as well as their strengths and weaknesses.

It is recommended that this chapter becomes your point of reference throughout the course. As you study the course, apply what you have learned here to new contexts and specific research studies. This way you will constantly exercise your skills and deepen your understanding.

3.1 Research methods in psychology: overview

All research methods used in psychology can be categorized as either quantitative or qualitative (see Figure 3.1).

3.1.1 Quantitative research

The aim of quantitative research is to arrive at numerically expressed laws that characterize the behaviour of large groups of individuals (universal laws).

Quantitative research operates with variables.

- A variable ("something that can take on various values") is any characteristic that is objectively registered and quantified.
- A construct is any theoretically defined variable—for example, violence, aggression, attraction, memory, attention, love, anxiety. Constructs cannot be directly observed.
- Operationalization of a construct means expressing it in terms of observable behaviour. For example, to operationalize anxiety you might look at the level of cortisol (the stress hormone) in the bloodstream.

Deeper dive

Note that a survey or questionnaire can be used in two ways.

First, it can be used as a research method in its own right. In this case, we are dealing with a quantitative descriptive study where the end goal is to obtain a description of the population in terms of the construct that is measured by the survey or questionnaire. For example, a study of prevalence rates aims to establish what percentage of the population can be said to experience severe depression, mild depression, no depression, and so on.

Second, it can be used as a data collection tool as part of a broader study. In this case, the research method would be either the experiment or a correlational study, and the survey or questionnaire would be used to measure a variable, but not as a research method in its own right. For example, suppose we investigate the relationship between aggressiveness and insecurity among high school students. We might measure their aggressiveness and insecurity with special questionnaires, but the point of the research is to investigate the correlation between these two variables. This would be a correlational study in which questionnaires are used as a tool of data collection.

There are three types of quantitative research—they differ in terms of the conclusions they allow us to make:

1. Experiment: allows making cause–effect conclusions of the type "variable A influences variable B".
2. Correlational study: allows making conclusions about relationships between variables (but not causation). For example, "variable A is related to variable B; as A increases, so does B".
3. Quantitative descriptive study: in descriptive studies relationships between variables are not investigated, and the variables are approached separately. An example would be a public opinion survey. We ask questions (e.g., "Do you support the current policies of the government?") and we are interested in the distribution of answers to this question, such as the percentage of people who said "yes".

3.1.2 Qualitative research

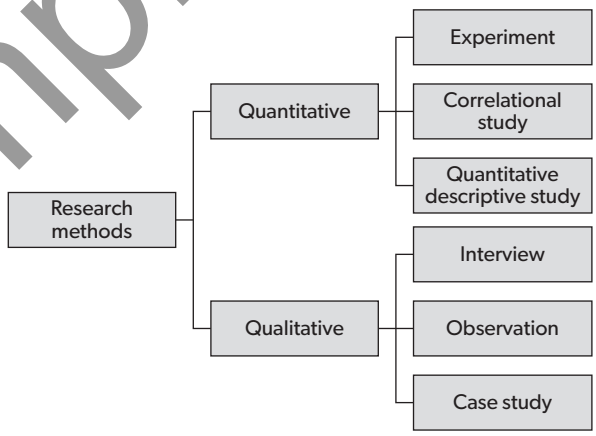
Qualitative research does not aim to arrive at universal laws. Instead, the focus is an in-depth study of a particular phenomenon.

Characteristics of qualitative research:

- It usually entails going beyond what can be objectively measured into the realm of human experiences, interpretations, and meanings.
- Data usually come in the form of texts, for example, interview transcripts or observational notes.
- Interpretation of data involves a degree of subjectivity, but analysis is deeper than we can usually achieve through quantitative approaches.

Qualitative research methods include:

1. Observation
2. Interview / focus group
3. Case study



▲ Figure 3.1 Research methods in psychology

3.2 Analysing research (methodological considerations)

3.2.1 Generalizability, sampling, credibility, and bias

Every study has strengths and limitations. There are several main parameters commonly used in making a judgement about the quality of a research study. They can be summarized as two pairs of parameters:

i. Generalizability and sampling

ii. Credibility and bias

Generalizability and sampling

Generalizability is the extent to which we can generalize the results of a research study beyond the study itself. There are three aspects of generalizability.

Aspect 1: Generalizing results from the participants who actually took part in the study (the sample) to other people.

- In quantitative methods (e.g., the experiment), the idea that results of a study can be generalized to larger groups of people is referred to as population validity. We believe that population validity is high if the sample is representative of the target population—that is, it reflects all the essential characteristics of the target population in miniature.
- In qualitative research, a similar idea is known as sample-to-population generalizability, or, in some sources, inferential generalizability. Sampling in qualitative research is not the same and so the idea of representativeness is not applicable.

Aspect 2: Generalizing results of the study to other situations and settings.

- In the experiment, the commonly used term is ecological validity. Ecological validity is the extent to which results of an experiment may be applied to real-life settings.
- In qualitative research, there is a similar idea known as transferability, or case-to-case generalizability. This is the extent to which we believe that the findings of a qualitative research study can be applied to another context.

Aspect 3: Generalizing from the data obtained in the study to the theoretical conclusion.

- In quantitative research (such as experiments, correlational studies and questionnaires), this idea is commonly referred to as construct validity. Construct validity is the extent to which we believe the study "measures" our hypothesized theoretical construct and not something else.
- In qualitative research, the term "construct validity" is not used, although a similar idea is referred to as theoretical generalizability.

Credibility and bias

Credibility is the degree to which the results of the study can be trusted to reflect reality. It is closely linked to bias. If there is an indication that potential sources of bias were, to the best of our knowledge and abilities, controlled or eliminated, the credibility of the study is believed to be high.

- In the experiment this idea is referred to as internal validity. Internal validity is high if we have reasons to believe that it is the manipulation of the independent variable, and not something else, that has led to the observable change in the dependent variable.
- In qualitative research the term "validity" is not used. However, the terms "credibility" or "trustworthiness" are used to refer to a similar idea. We believe that the results of a qualitative study are credible if we think that the way we formulated conclusions reflects participants' actual experiences.

3.2.2 Overview of methodological considerations

Table 3.1 provides an overview of the main concepts used to characterize sampling, generalizability, credibility and bias in experimental as opposed to qualitative research.

Overarching idea	Quantitative research (the experiment)	Qualitative research (interview, observation)
Generalizing results to a larger group of people	Population validity*	Sample-to-population generalizability / inferential generalizability
Generalizing results to real-life settings	Ecological validity*	Case-to-case generalizability / transferability
Generalizing results from the measurement to the theoretical construct	Construct validity	Theoretical generalizability
The degree to which results of the study reflect reality	Internal validity	Credibility / trustworthiness

Together population validity and ecological validity of the experiment are also known as external validity

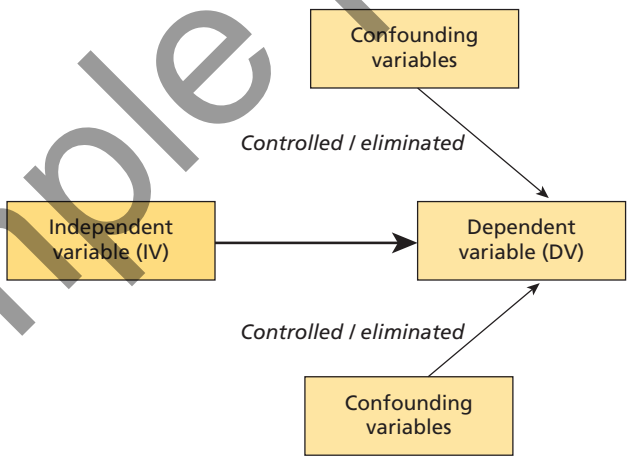
▲ Table 3.1 Overview of methodological considerations

3.3 The experiment in quantitative research

3.3.1 Variables in the experiment

The simplest experiment includes one independent variable (IV) and one dependent variable (DV), while the other potentially important variables are controlled (see Figure 3.2).

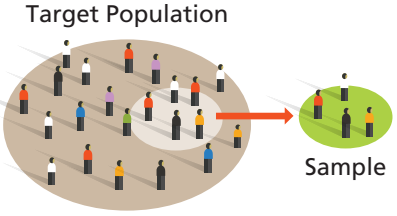
The IV is manipulated by the experimenter. The DV changes because of this manipulation. All the other variables that can interfere in the relationship between the IV and the DV are called confounding variables. Confounding variables need to be controlled (i.e., kept constant).



▲ Figure 3.2 Variables in an experiment

3.3.2 Sampling in quantitative research

We need to be sure that results of quantitative research can be generalized from the sample to the target population. For this to be possible, the sample must be representative of the target population, i.e., it must reflect all the essential characteristics of the target population (see Figure 3.3).



▲ Figure 3.3 Sample and target population

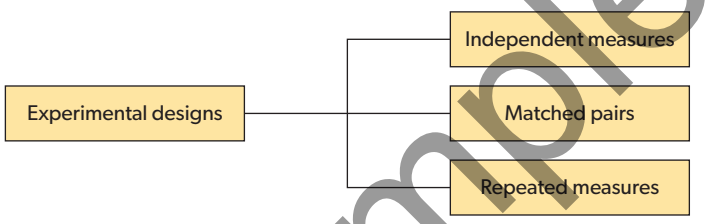
Several sampling techniques can be used in quantitative research. The choice depends on the aim of the research, available resources and the nature of the target population (see Table 3.2).

Sampling technique	Explanation	Advantages	Disadvantages
Random sampling	Create a list of all members of the target population and randomly select a subset.	If the sample size is sufficient, researchers may be certain that even unexpected characteristics are fairly represented in the sample.	With large populations, it is practically impossible to carry out truly random sampling.
Stratified sampling	First determine the proportions of essential characteristics (e.g., age, occupation, language) in the population. Then recruit participants in a way that keeps the same proportions in the sample.	Allows researchers to control representativeness of some key characteristics without relying on chance.	Requires more knowledge about characteristics of the target population; harder to implement.
Convenience (opportunity) sampling	Recruiting participants that are readily available (e.g., undergraduate psychology students).	Useful when financial resources are limited. Also useful when generalization of findings is not the primary purpose of the study.	Generalization from opportunity samples is very limited because of the sampling bias.
Self-selected sampling	Recruiting volunteers, for example, through newspaper advertisements. Anyone who wants to participate is included in the sample.	A quick and easy method to recruit participants while simultaneously having wide coverage (e.g., many different types of people read newspapers).	Representativeness and generalization are limited. A typical volunteer is different from the average participant from a larger population (e.g., they are more motivated).

▲ Table 3.2 Sampling techniques in qualitative research

3.3.3 Experimental designs

There are three types of experimental design, depending on how the independent variable is manipulated (see Figure 3.4).



▲ Figure 3.4 Types of experimental design

Independent measures design

This involves random allocation of participants into groups and a comparison between these groups (see Figure 3.5).

When the group sizes are sufficiently large and allocation is random, chances are that groups will be equivalent. This allows us to assume that whatever difference we observe at the end of the experiment must have been caused by our experimental manipulation.

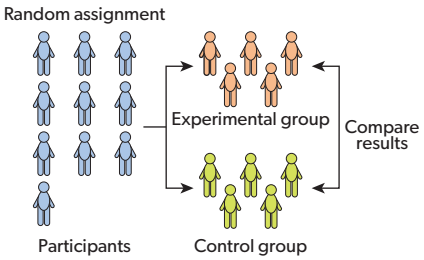
Matched pairs design

This is like independent measures, but instead of completely random allocation, researchers use matching to form the groups (see Figure 3.6).

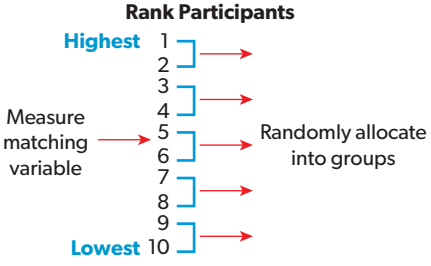
Suppose you want to use age as the matching variable. You will take the two oldest participants in your sample and allocate them randomly to one of two groups (experimental or control), then take the next two oldest participants, and so on. The result is two groups believed to be equivalent in terms of age but also formed randomly in terms of all other characteristics.

Assessment tip

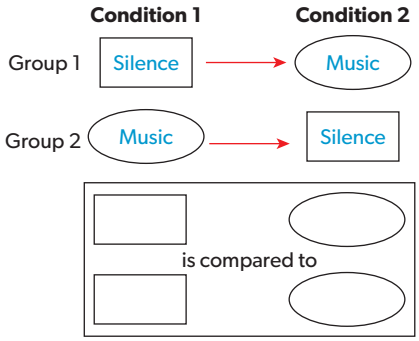
Be mindful of the difference between random sampling (selecting the sample from the target population) and random group allocation (dividing your sample into groups). It is possible to have random group allocation in non-random samples, and vice versa.



▲ Figure 3.5 Independent measures design



▲ Figure 3.6 Matched pairs design



▲ Figure 3.7 Repeated measures design with counterbalancing

Repeated measures design

This involves exposing the same group of participants to two (or more) conditions, and then comparing the results of the conditions. This way participants are compared to themselves (see Figure 3.7).

Suppose your aim is to investigate the effect of classical music on learning. You ask your participants to learn a list of words for 10 minutes in silence. You then ask the same participants to learn a different list of words for another 10 minutes, but this time with classical music playing in the background. You compare results from the first and the second trial.

The problem with repeated measures designs is that they are vulnerable to order effects: results may be different depending on which condition comes first. To minimize this problem, counterbalancing is used: participants are divided into two groups that only differ in the order of conditions (Group 1: Condition A–Condition B; Group 2: Condition B–Condition A).

3.3.4 Types of validity

The quality of experiments is characterized by their construct, internal, and external validity (see Figure 3.8).

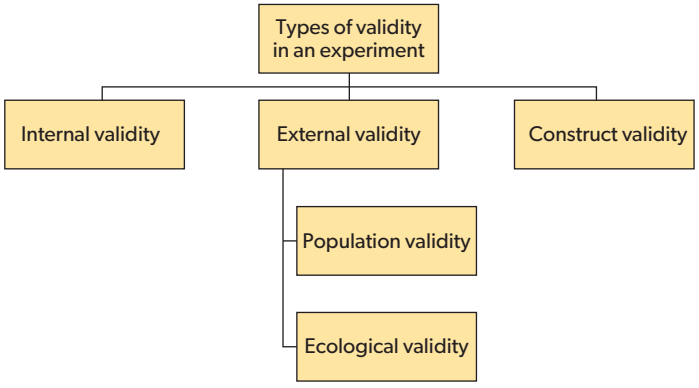
Construct validity is a characteristic of the quality of operationalizations. For example, anger is a construct, and a response to the question "How angry are you at the moment?" on a scale from 1 to 5 is its operationalization. Shifting from an operationalization to a construct is a conceptual move. An experiment's construct validity is strong if this move is defensible and if the operationalization sufficiently covers the construct. In this sense, construct validity relates to generalizability—it characterizes generalizability of findings to the theory.

External validity is a characteristic of the generalizability of findings to other people and other situations. There are two types of external validity:

1. **Population validity** refers to the extent to which findings can be generalized from the sample to the target population. It depends on how representative the sample is.
2. **Ecological validity** refers to the extent to which findings can be generalized from the experiment to other settings or situations. It depends on how artificial the experimental procedure is. In laboratory experiments, participants often find themselves in situations that do not normally occur in their daily lives, and this can change their behaviour, making it less natural.

Internal validity is a characteristic of the methodological quality of an experiment—its credibility. Internal validity is high when confounding variables have been controlled and we are quite certain that it was the change in the IV (not something else) that caused the change in the DV. Internal validity links directly to bias—the less bias, the higher the internal validity of the experiment.

Usually there is an inverse relationship between internal validity and ecological validity—as one increases, the other decreases. For example, in an attempt to increase internal validity, we might conduct the experiment in well-controlled conditions in a laboratory. However, this would also imply that the situation is now artificial and that participants will not necessarily behave the same way as they do in real life.



▲ Figure 3.8 Types of validity in experiments

3.3.5 Bias in experimental research: threats to internal validity

Bias in experimental research comes in the form of confounding variables that can reduce internal validity. Here are several common sources of threat to internal validity (based on Campbell, 1957). See Table 3.3.

Threat to internal validity	Explanation	How it can be counteracted
Selection	For some reason groups are not entirely equivalent at the start of the experiment (like comparing apples and oranges).	Random allocation into groups; sufficiently large group sizes.
Maturation	The natural changes that participants go through during the experiment, such as fatigue or natural growth.	Have a control group. If maturation is the same in both groups, the comparison will not be affected.
Testing effect	The first measurement of the DV may affect the second (and subsequent) measurements.	In repeated measures designs, counterbalancing must be used.
Experimental mortality	Occurs when some participants drop out of the experiment and the rate of dropping out is not the same in all groups.	Design experimental conditions so that participants do not feel discomfort causing them to withdraw.
Demand characteristics	Occurs when participants understand the true aim of the experiment and alter their behaviour because of that.	Use deception to conceal the true aim of the study (however, ethical considerations would then arise).
Experimenter bias	Occurs when the researcher unintentionally influences participants' behaviour and the results of the study.	Use a double-blind design: neither the participants nor the experimenter knows who has been assigned to what condition.

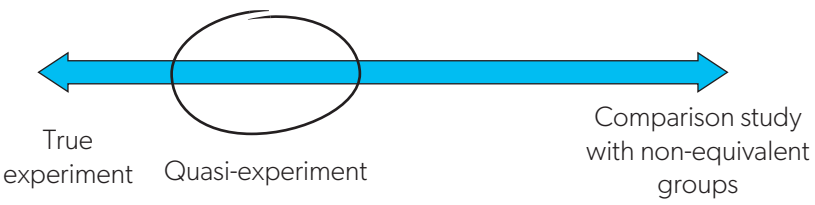
▲ Table 3.3 Threats to internal validity in experimental research

3.3.6 True experiments versus quasi-experiments versus non-experiments

The key feature of a true experiment is that we assume the groups are equivalent in every way except for the IV. To make this possible, we rely on random allocation of participants into groups. This enables us to make cause–effect inferences.

The opposite situation (a non-experiment) is when we use pre-existing groups. For example, we compare male participants to female participants, Chinese to Canadian, teenagers to adults. All these are examples of non-equivalent groups. Such groups are different from each other in multiple ways. If we discover that a measurement (such as memory) is different in the two groups, we cannot be sure what exactly this difference should be attributed to. We cannot make cause–effect inferences.

There is a grey area in between where we have studies with group allocation that was controlled, but which are not entirely random. Such studies are closer to true experiments, but they lack some features of a true experiment.



▲ Figure 3.9 "Grey area"—quasi-experiments

For example, Sharot et al. (2007) compared two groups of participants who witnessed the 9/11 attacks in New York City in 2001: those who were close to the attack site in Downtown Manhattan when it happened, and those who were in Midtown, a few miles away. One may argue that it was a matter of chance in what part of the city a participant was when the attack took place. On the other hand, Sharot did not allocate participants into the two conditions randomly herself, so it is still possible that there are pre-existing differences that can affect the results of the study.

There are two ways of looking at these studies residing in the grey area (see Figure 3.9). There is no consensus among researchers on which is the best approach, leading to some debate on this issue:

1. We could strictly draw a line between true experiments and non-experiments by saying that anything that does not have all features of a true experiment is a non-experiment. Any study that did not use random allocation into groups or conditions will be considered a non-experiment. Cause–effect inferences from such studies will not be allowed.
2. We could use a softer approach and allow some of the studies from the "grey area" to be interpreted in terms of cause and effect, albeit with caution. It is within this approach that the term quasi-experiment was coined by Cook and Campbell (1979). "Quasi" is a prefix meaning "almost".

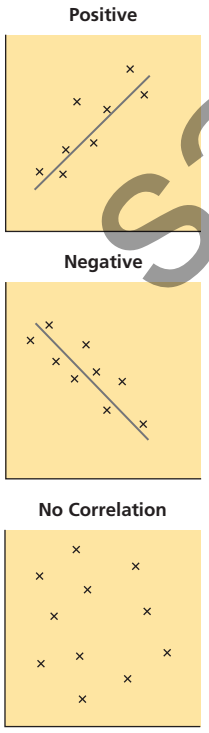
3.3.7 Natural experiments

Natural experiments can be seen as a special type of a quasi-experiment. In natural experiments, the experimental manipulation occurs naturally and has not been caused by a researcher, although there are reasons to believe that group allocation (i.e., people who experienced the natural event versus those who did not) is fairly random.

3.3.8 Laboratory experiments and field experiments

Laboratory experiments involve inviting participants to a specially designated venue where the study is conducted according to a standard script, in well-controlled conditions. Their advantage is higher internal validity. However, the cost of this is usually lower ecological validity. Well-controlled environments are also artificial, and behaviour of people in an artificial situation is not necessarily the same as it would be in similar conditions in real life.

Field experiments are conducted in a real-life setting. The researcher manipulates the IV, but since participants are in their natural setting many extraneous variables cannot be controlled.



▲ Figure 3.10 Examples of correlations

3.4 Correlational studies in quantitative research

3.4.1 What is a correlation?

A correlation is a measure of the linear relationship between two variables (see Figure 3.10). A correlation coefficient can vary from −1 to +1.

- A negative correlation means that there is an inverse relationship between two variables: the higher A, the lower B.
- A positive correlation means a direct relationship: the higher A, the higher B.
- A correlation close to zero means that there is no relationship between the two variables.