

OXFORD IB DIPLOMA PROGRAMME



2ND EDITION

# PSYCHOLOGY

COURSE COMPANION

Alexey Popov  
Lee Parker  
Darren Seath

OXFORD

Great Clarendon Street, Oxford, OX2 6DP, United Kingdom

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

© Oxford University Press 2017

The moral rights of the authors have been asserted.

First published in 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by licence or under terms agreed with the appropriate reprographics rights organization. Enquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form and you must impose this same condition on any acquirer.

British Library Cataloguing in Publication Data  
Data available

978-0-19-839811-0

10 9 8 7 6 5 4 3 2 1

Paper used in the production of this book is a natural, recyclable product made from wood grown in sustainable forests. The manufacturing process conforms to the environmental regulations of the country of origin.

Printed in United Kingdom

### Acknowledgements

The author and publisher are grateful for permission to reprint extracts from the following copyright material:

**Ajzen, Icek:** 'The theory of planned behaviour is alive and well, and not ready to retire: a commentary on Sniehotta, Presseau, and Araújo-Soare' from *Health Psychology Review*, Vol. 9, Issue 2, published by Taylor & Francis, 2015, reprinted by permission.

**Allport, Gordon W:** *The Nature of Prejudice*, Copyright 1954 by Perseus Books Publishing L.L.C., reprinted by permission of Perseus Books, a member of the Perseus Books Group, permission conveyed through Copyright Clearance Center, Inc.

**Baddeley, Alan:** 'Exploring the central executive', in *The Quarterly Journal of Experimental Psychology*, 1996, Vol. 49A, (1), pp. 5-28, reprinted by permission of the publisher (Taylor & Francis Ltd., [www.tandfonline.com](http://www.tandfonline.com)).

**Baron, Jonathan:** *Thinking and Deciding* (fourth edition), Cambridge University Press, 2008.

**Bower, GH, Black, JB and Turner, TJ:** 'Scripts in memory for text', in *Cognitive Psychology*, 1979, Vol. 11, pp. 177-220, reprinted by permission of Elsevier.

**Bransford, JD & Johnson, MK:** p. 719, in *Visual Information Processing*, 1973, W G Chase, New York, Academic Press, reprinted by permission of Elsevier.

**Burbank, V. K:** 'Fighting Women. Anger and Aggression in Aboriginal Australia', Berkeley: University of California Press, (1994), reprinted by permission of University of California Press Books.

**Center for Science in the Public Interest:** report 'Carbonating the World - The Marketing and Health Impact of Sugar Drinks in Low- and Middle-income Countries', reprinted by permission.

**Cutler, WB, Friedmann, E and McCoy, NL:** 'Pheromonal influences on sociosexual behavior in men' by, in *Archives of Sexual Behavior*, 1998, Vol. 21, number 1, pp. 1-13, reprinted by permission of Springer.

**Hamilton, M:** 'A rating scale for depression', 1960, in *Journal of Neurology, Neurosurgery, and Psychiatry*, Vol. 23, pp. 56-62, reprinted by permission of BMJ Publishing Group Ltd.

**Holmes, Thomas H and Rahe, Richard H:** abridged version of table 'The Social Readjustment Rating Scale' in *Journal of Psychosomatic Research*, 1967, Vol. 11, Issue 2, pp. 213-218, Copyright 1967, reprinted by permission of Elsevier.

**Johnson, D H and Johnson F P:** *Joining Together: Group Theory and Group Skills*, 1st Ed., © 1975, reprinted by permission of Pearson Education, Inc., New York, New York.

**International Baccalaureate Organization:** various extracts from *IB Diploma Programme Psychology Guide* © International Baccalaureate Organization, 2017.

**Kabat-Zinn, J:** 'Some reflections on the origins of MBSR, skilful means, and the trouble with maps' in *Contemporary Buddhism*, 2011, Vol. 12, 1, pp. 281-306, reprinted by permission of the publisher (Taylor & Francis Ltd), <http://www.tandfonline.com>.

**Kahneman, D and Tversky, A:** 'The framing of decisions and the psychology of choice', in *Science*, Vol. 211, pp. 453-458, reprinted by permission of AAAS.

**Selye, H:** 'The General-Adaptation-Syndrome' in *Annual Review of Medicine*, 1951, Vol. 2, pp. 327-342, © by Annual Reviews, <http://www.annualreviews.org>, reprinted by permission of Annual Reviews.

**Skinner, BF:** 'Superstition' in the pigeon', in *Journal of Experimental Psychology*, 1948, Vol. 38 (2), pp. 168-172, this material is in the public domain.

**Talarico, JM and Rubin, DC:** 'Confidence, not consistency, characterizes flashbulb memories', 2003, in *Psychological Science*, Vol 14, 5, SAGE Journals.

**World Health Organization:** report 'World Health Statistics 2016'

[http://www.who.int/gho/publications/world\\_health\\_statistics/2016/en/](http://www.who.int/gho/publications/world_health_statistics/2016/en/), copyright WHO 2016, reprinted by permission.

**World Health Organization:** 'Obesity and overweight' - Fact sheet N°311 - updated June 2016, <http://www.who.int/mediacentre/factsheets/fs311/en/>, copyright WHO 2016, reprinted by permission.

World Health Organization, Health Topics, Health Promotion, [http://www.who.int/topics/health\\_promotion/en/](http://www.who.int/topics/health_promotion/en/), copyright WHO, reprinted by permission.

**World Health Organization:** 'Fiscal Policies for Diet and Prevention of Non-communicable Diseases', copyright WHO, 2016, reprinted by permission.

Although we have made every effort to trace and contact all copyright holders before publication, this has not been possible in all cases. If notified, the publisher will rectify any errors or omissions at the earliest opportunity.

Links to third party websites are provided by Oxford in good faith and for information only. Oxford disclaims any responsibility for the materials contained in any third party website referenced in this book.

Photo and artwork permissions are continued on the last page

## Course Companion definition

The IB Diploma Programme Course Companions are resource materials designed to support students throughout their two-year Diploma Programme course of study in a particular subject. They will help students gain an understanding of what is expected from the study of an IB Diploma Programme subject while presenting content in a way that illustrates the purpose and aims of the IB. They reflect the philosophy and approach of the IB and encourage a deep understanding of each subject by making connections to wider issues and providing opportunities for critical thinking.

The books mirror the IB philosophy of viewing the curriculum in terms of a whole-course approach; the use of a wide range of resources; international-mindedness; the IB learner profile and the IB Diploma Programme core

## IB mission statement

The International Baccalaureate aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.

To this end the IB works with schools, governments and international organizations to

requirements; theory of knowledge; the extended essay; and creativity, action, service (CAS).

Each book can be used in conjunction with other materials and indeed, students of the IB are required and encouraged to draw conclusions from a variety of resources. Suggestions for additional and further reading are given in each book and suggestions for how to extend research are provided.

In addition, the Course Companions provide advice and guidance on the specific course assessment requirements and on academic honesty protocol. They are distinctive and authoritative without being prescriptive.

develop challenging programmes of international education and rigorous assessment.

These programmes encourage students across the world to become active, compassionate, and lifelong learners who understand that other people, with their differences, can also be right.

## The IB learner profile

The aim of all IB programmes is to develop internationally minded people who, recognizing their common humanity and shared guardianship of the planet, help to create a better and more peaceful world. IB learners strive to be:

**Inquirers** They develop their natural curiosity. They acquire the skills necessary to conduct inquiry and research and show independence in learning. They actively enjoy learning and this love of learning will be sustained throughout their lives.

**Knowledgeable** They explore concepts, ideas, and issues that have local and global significance. In so doing, they acquire in-depth knowledge and develop understanding across a broad and balanced range of disciplines.

**Thinkers** They exercise initiative in applying thinking skills critically and creatively to recognize and approach complex problems, and make reasoned, ethical decisions.

**Communicators** They understand and express ideas and information confidently and creatively in more than one language and in a variety of modes of communication. They work effectively and willingly in collaboration with others.

**Principled** They act with integrity and honesty, with a strong sense of fairness, justice, and respect for the dignity of the individual, groups, and communities. They take responsibility for their own actions and the consequences that accompany them.



**Open-minded** They understand and appreciate their own cultures and personal histories, and are open to the perspectives, values, and traditions of other individuals and communities. They are accustomed to seeking and evaluating a range of points of view, and are willing to grow from the experience.

**Caring** They show empathy, compassion, and respect towards the needs and feelings of others. They have a personal commitment to service, and act to make a positive difference to the lives of others and to the environment.

**Risk-takers** They approach unfamiliar situations and uncertainty with courage and forethought,

and have the independence of spirit to explore new roles, ideas, and strategies. They are brave and articulate in defending their beliefs.

**Balanced** They understand the importance of intellectual, physical, and emotional balance to achieve personal well-being for themselves and others.

**Reflective** They give thoughtful consideration to their own learning and experience. They are able to assess and understand their strengths and limitations in order to support their learning and personal development.

## A note on academic honesty

It is of vital importance to acknowledge and appropriately credit the owners of information when that information is used in your work. After all, owners of ideas (intellectual property) have property rights. To have an authentic piece of work, it must be based on your individual and original ideas with the work of others fully acknowledged. Therefore, all assignments, written or oral, completed for assessment must use your own language and expression. Where sources are used or referred to, whether in the form of direct quotation or paraphrase, such sources must be appropriately acknowledged.

### How do I acknowledge the work of others?

The way that you acknowledge that you have used the ideas of other people is through the use of footnotes and bibliographies.

**Footnotes** (placed at the bottom of a page) or **endnotes** (placed at the end of a document) are to be provided when you quote or paraphrase from another document, or closely summarize the information provided in another document. You do not need to provide a footnote for information that is part of a 'body of knowledge'. That is, definitions do not need to be footnoted as they are part of the assumed knowledge.

**Bibliographies** should include a formal list of the resources that you used in your work. The listing should include all resources, including books, magazines, newspaper articles, Internet based resources, CDs and works of art. 'Formal' means that you should use one of the several accepted forms of presentation. You must provide full information as

to how a reader or viewer of your work can find the same information. A bibliography is compulsory in the extended essay.

### What constitutes misconduct?

**Misconduct** is behaviour that results in, or may result in, you or any student gaining an unfair advantage in one or more assessment component. Misconduct includes plagiarism and collusion.

**Plagiarism** is defined as the representation of the ideas or work of another person as your own. The following are some of the ways to avoid plagiarism:

- Words and ideas of another person used to support one's arguments must be acknowledged.
- Passages that are quoted verbatim must be enclosed within quotation marks and acknowledged.
- CD-ROMs, email messages, web sites on the Internet, and any other electronic media must be treated in the same way as books and journals.
- The sources of all photographs, maps, illustrations, computer programs, data, graphs, audio-visual, and similar material must be acknowledged if they are not your own work.
- Works of art, whether music, film, dance, theatre arts, or visual arts, and where the creative use of a part of a work takes place, must be acknowledged.

**Collusion** is defined as supporting malpractice by another student. This includes:

- allowing your work to be copied or submitted for assessment by another student

- duplicating work for different assessment components and/or diploma requirements.

Other forms of misconduct include any action that gives you an unfair advantage or affects the results

## Contents

Introduction ..... vi

### Unit 1 Research methodology

Research in psychology ..... 2  
 Quantitative research: the experiment ..... 9  
 Quantitative research: correlational studies..... 20  
 Qualitative research ..... 24  
 Qualitative research methods..... 31  
 Ethics in psychological research..... 37

### Unit 2 Biological approach to behaviour

Localization ..... 46  
 Neuroplasticity ..... 56  
 Neurotransmitters and behaviour ..... 64  
 Techniques used to study the brain in relation to behaviour ..... 72  
 Hormones and behaviour ..... 78  
 Pheromones and behaviour ..... 87  
 Genes and behaviour; genetic similarities..... 94  
 Evolutionary explanations for behaviour ..... 108  
 The role of animal research in understanding human behaviour (HL only)..... 115

### Unit 3 Cognitive approach to behaviour

Concepts and principles of the cognitive approach to behaviour ..... 123  
 Models of memory ..... 131  
 Schema theory ..... 141  
 Thinking and decision-making..... 149  
 Reliability of cognitive processes:  
 reconstructive memory ..... 160  
 Biases in thinking and decision-making ..... 168  
 Emotion and cognition..... 180  
 Cognitive processing in the digital world (HL only) ..... 188

### Unit 4 Sociocultural approach to behaviour

Cultural influences on individual attitudes, identities and behaviour—culture and cultural norms ..... 202  
 Cultural origins of behaviour and cognition—cultural dimensions ..... 208  
 The individual and the group—social cognitive theory ..... 220  
 The individual and the group—social identity theory and stereotyping ..... 229

of another student. Examples include, taking unauthorized material into an examination room, misconduct during an examination, and falsifying a CAS record.

The influence of globalization on individual behaviour (HL only)..... 250

### Unit 5 Abnormal psychology

Normality versus abnormality..... 257  
 Classification systems ..... 262  
 Prevalence rates and disorders ..... 266  
 Validity and reliability of diagnosis ..... 269  
 The role of clinical biases in diagnosis ..... 280  
 Explanations for disorders—biological explanations for depression ..... 288  
 Explanations for disorders—cognitive and sociocultural explanations for depression ..... 295  
 Assessing the effectiveness of treatment..... 300  
 Biological treatment of depression..... 306  
 Psychological treatment of depression ..... 313  
 The role of culture in treatment ..... 319

### Unit 6 Health psychology

Determinants of health ..... 327  
 Health problem 1: obesity and being overweight ..... 331  
 Social cognitive explanations of health problems ..... 339  
 Promoting health ..... 344  
 Health problem 2: stress..... 350

### Unit 7 Psychology of human relationships

Personal relationships ..... 370  
 Group dynamics ..... 386  
 Prejudice and discrimination..... 394  
 Origins of conflict and conflict resolution..... 400  
 Social responsibility..... 404

### Unit 8 Developmental psychology

Brain development..... 422  
 Cognitive development ..... 430  
 Development of empathy and theory of mind ..... 437  
 Gender identity and social roles..... 446  
 Attachment..... 455  
 Role of peers and play..... 464  
 Childhood trauma and resilience ..... 473  
 Effects of poverty on relationships..... 481

### Unit 9 Internal assessment

Overview of the requirements for internal assessment ..... 488

Planning the investigation .....	490
Writing the introduction .....	493
Writing the exploration.....	495
Conducting the analysis .....	497
Writing the evaluation .....	510
References and appendices .....	512

The extended essay .....	521
Other resources .....	528
Bibliography .....	529
Index .....	564

**Unit 10 The IB curriculum:  
a conceptual model**

Examination requirements .....	516
--------------------------------	-----

## Introduction

This book is a Course Companion for students of psychology in the International Baccalaureate Diploma Programme at higher and standard levels.

It is designed to be used extensively both in class and at home. We have tried to provide deep coverage of all topics in the syllabus, focusing on a variety of arguments supported by both classic and contemporary research. Your job is to study the material along with the discussions and activities that you have in class, take notes, make mind maps and use other techniques to “compress” the information and understand the topics. When you eat, you chew the food before swallowing it, which helps you digest the food better. This Course Companion is no different: if you want to digest it, you need to chew it (this is a metaphor, so please do not literally chew the book).

To help you, the book has a number of features that you can use to enhance your learning skills.

- Inquiry questions at the start of every section will encourage you to think about problems that do not have an easy solution. Take a stance, but be ready to change it as you discover new knowledge.
- Material provided in the text will equip you with a range of arguments and evaluation points to deepen the inquiry questions and uncover their hidden dimensions.
- These arguments will be supported with empirical research, because knowledge in psychology is procedural: in order to understand it fully, you need to know how it was obtained.
- “What you will learn in this section” boxes will help you summarize the key points.
- “This section also links to” will support you in making links between topics—this is an

important skill because human behaviour is complex and should be studied in all its aspects.

- “Psychology in real life” features will apply the concepts discussed in the text to some real-life scenarios, to help you see the vast practical applications.
- “ATL skills” will suggest a number of questions and activities for you and your classmates to develop your learning skills further, and help you become better researchers and communicators.
- TOK links will challenge you to see knowledge concepts behind key psychological terms, allowing you to compare psychology meaningfully to other disciplines.
- Exercises and links to external materials (such as videos and research papers) will take you beyond this book while at the same time staying focused on the topics relevant to the syllabus.

There are two overarching themes in everything that we are discussing in this book: research and ethics. Research is important because every claim in psychology has a history of discovery behind it. Ethics is important because the focus of research in psychology is living beings—human and non-human animals—so research needs to be done responsibly.

Your teacher will provide you with information from the IB psychology subject guide: aims of the course, list of topics that you need to know, assessment requirements and criteria. Remember to refer to this information so that you clearly understand what is expected of you at all times.

Psychology is a journey full of exciting discoveries—but no spoilers—you will see for yourself.

*Alexey Popov, Lee Parker, Darren Seath*

# RESEARCH METHODOLOGY

## Topics

- Introduction: Research in psychology
- Quantitative research: the experiment
- Quantitative research: correlational studies
- Qualitative research
- Qualitative research methods
- Ethics in psychological research

## Introduction

This unit deals with research methods used in psychology. In any discipline, knowledge of research methods greatly increases our ability to understand a topic. Psychology is not an exception. The ability to evaluate psychological knowledge critically on the basis of how it was obtained is essential to avoid misconceptions.

Speaking of misconceptions, there are plenty of them in this field. Psychology is a popular discipline which makes it vulnerable to numerous popular interpretations. So it is important to clearly understand what psychology is and what it is not.

Knowledge of methods also allows you to see the whole research process clearly, with all its strengths and limitations. When studying the material in this chapter, you will no longer take statements like “British psychologists have discovered ...” at face value. You will read between the lines and understand what was done by the “British psychologists” and to what extent their inferences are justified.

Psychology is a special discipline. On the one hand, it is scientific, which means that psychologists, just like physicists or chemists, rigorously test hypotheses and eliminate competing explanations in an attempt to achieve objective knowledge. On the other hand, unlike natural sciences that study “nature”, psychology studies humans, inherently subjective creatures. So psychology is an attempt to study the subjective (for example,

the mind) objectively. Not an easy task, if you think about it.

This unit may seem a little abstract to you but it builds an important foundation for the understanding of the material in all other units. Applying the knowledge and skills related to research methodology, you will be able to critically evaluate knowledge in specific areas of psychology and arrive at balanced conclusions, avoiding misconceptions and unjustified generalizations. We will keep referring you back to this unit so that you can apply and reinforce the concepts related to research methodology.

We start by discussing the definition of psychology, what it is and what it is not. Then we introduce two broad groups of research methods: quantitative and qualitative. These two groups of methods differ dramatically in their rationale and objectives, but at the same time can be combined to complement each other in a holistic investigation. Following this, we discuss four overarching concepts that apply to both quantitative and qualitative research: sampling, credibility, generalizability and bias. Next, we look at the application of these concepts separately in quantitative (experiments, correlational studies) and qualitative research. Finally, any discipline that involves research with living beings needs to adhere to the principles of ethics. We discuss ethical considerations in psychological research.



# 1 Research in psychology

## Inquiry questions

- What is scientific psychology?
- How can we tell if a research study is credible?
- How can we study subjective phenomena objectively?
- How is correlation different from causation?
- How is quantitative research different from qualitative research?

## What you will learn in this section

- What is psychology?
  - Psychology is the scientific study of behaviour and mental processes
  - Science and non-science
  - Behaviour and mental processes
  - A study of non-human animals
  - What IB psychology is not
- Research methodology: quantitative and qualitative methods
  - Qualitative versus quantitative comparison
  - Types of quantitative research: experimental, correlational, descriptive
  - Types of qualitative research
- Sampling, credibility, generalizability and bias in research: an overview

## What is psychology?

“Psychology is the scientific study of behaviour and mental processes.” This is the definition we are going to use throughout this book. Although it is quite a short definition, there are a lot of implications in it. Let’s try and uncover them one by one.

*Psychology is the scientific study...* This part of the definition excludes such areas as pop psychology, that is, simple and appealing explanations that are not backed up by empirical evidence. What makes a theory or a study scientific, or where is the line between science and non-science? This is largely a TOK question and you will return to it throughout the book, but here are some major points.

- It should be supported by empirical evidence and be based on this evidence.
- It should be falsifiable, that is, it should be possible for the theory or study to be proven wrong.

- There should be a history of independent attempts to test the theory or replicate the study.

## TOK

Science versus non-science demarcation is one of the key topics in TOK. The following concepts are important in the discussion of demarcation criteria:

- empirical evidence
- falsification/falsifiability
- replication.

While reading this unit, take a note of examples that illustrate these three concepts.

Think of other similar examples from such areas of knowledge as human sciences, natural sciences and mathematics.



## Exercise

Look at the following research questions and pick one that you find interesting:

1. Do children who watch more violent TV shows become more violent?
2. Does extrasensory perception exist?
3. Are women attracted to men by the smell of their body?
4. Is abuse experienced differently in heterosexual and gay relationships?
5. Are breathing exercises effective for reducing test anxiety?
6. What emotions do people experience when watching horror movies in a cinema?
7. Are people in arranged marriages happier than people who married by choice?

If you were to conduct a research study to answer the question that you picked, how would you go about it? Think about details such as who your participants would be, what they would be required to do, how you would measure results and how you would ensure that the results are believable.



▲ Figure 1.1 Wilhelm von Osten and Clever Hans

In the early twentieth century, under the influence of Charles Darwin's theory of evolution, the public was very interested in animal intelligence: if humans evolved from animals, animals must be at least partially intelligent, so what exactly are they capable of? The case of Clever Hans sparked a lot of interest. Hans was a horse. Its owner Wilhelm von Osten, a mathematics teacher, claimed that

he had taught Hans to solve arithmetic problems (addition, subtraction, multiplication, division, fractions), read, spell and understand some German. Questions could be asked verbally or in writing, and Hans would respond by tapping his hoof a certain number of times. Von Osten exhibited the horse frequently and gained a lot of public attention. A special committee was formed in Germany (called the Hans Commission). They ran a series of tests and concluded that the performance was not a fraud. So Hans's abilities were officially recognized as phenomenal!

However, another independent investigation carried out later by Oskar Pfungst, a psychologist, yielded different results. It demonstrated that Hans could not actually perform mental operations such as multiplication, but the horse was very responsive to clues that were provided by unsuspecting humans. To arrive at these conclusions, Pfungst successively tested a number of alternative hypotheses.

1. What if spectators give the horse hints or clues? He tested the horse and the questioner in the absence of spectators, but the horse continued to solve tasks correctly anyway.
2. What if von Osten himself gives the horse some clues? Another questioner was used during several trials, but the horse's performance did not worsen.
3. What if something in the questioner gives the correct answer away and the horse can feel that? Blinders were used to test this hypothesis. It turned out that when Hans was wearing blinders responses (the number of hoof taps) were incorrect most of the time. So, it was something in the questioner after all.
4. Did the questioners consciously let the horse know the correct answer, though? Additional trials were organized so that the questioner either knew or did not know the answer to the questions. It turned out that Clever Hans could only answer the questions correctly when the questioner knew the answer in advance.

This changed the focus of research from the horse to the questioner. When Pfungst carried out his observations, it was concluded that questioners who knew the answers had a tendency to become more tense as the hoof tapping approached the correct answer which would be reflected in their posture and facial expressions without them

realizing it. This was probably the clue that the horse was using. This makes sense evolutionarily, as detection of small postural changes is important as a survival skill for horses in the wild. Clever Hans certainly was clever, but the nature of his abilities was not mathematical (Goodwin, 2010)!

### ATL skills: Thinking

How does Pfungst's investigation illustrate the concepts of empirical evidence, falsification and replication?

Von Osten himself, however, was never convinced of Pfungst's findings and he continued to exhibit the horse throughout Germany, gaining as much popularity as before. Nonetheless, scientifically, this was one of the starting points for designing rigorous experimental methodology in psychology and other human sciences. It was recognized that experiments, if not carefully controlled, could produce **artifacts**—results that are associated with the effect of unforeseen factors.

This whole story shows how claims can and should be tested scientifically, that is, by conducting a systematic evidence-based investigation that puts forward one hypothesis after another and tests them in a rigorous fashion. Note also how the whole investigation attempted to falsify the existing theory rather than support it.

... *study of behaviour and mental processes.* A scientific investigation requires an empirical approach to research, that is, relying on observation as a means of data collection. On the other hand, psychology (which comes from the Greek *psyche* = soul and *logos* = study, "the study of the soul") concerns itself with a wealth of phenomena, many of which are not directly observable. The first step in solving this dilemma is to identify something that can be observed directly. That's **behaviour**. Behaviour is everything that can be registered by an independent observer: it includes overt actions as well as gestures, facial expressions, verbal responses, endocrine reactions and so on. What stays "behind the scene" are the **mental processes** such as attention, perception, memory and thinking. We cannot observe them directly (which led some psychologists to say that they represent a "black box" and cannot be studied

scientifically), but we can observe the indirect effects mental processes have on one's behaviour. So, we can infer something about the mental world as well.

### ATL skills: Thinking

Brainstorm some behavioural indicators of the following:

- attention
- anxiety
- embarrassment.

To what extent do you think is it possible to use behavioural indicators to infer these "internal" phenomena? Would the inference be reliable?

Throughout this book we will use the term "behaviour" to refer to external, observable manifestations while the term "mental processes" will be used to denote internal patterns of information processing. However, you need to be aware of the fact that the term "behaviour" is often used in a more general sense, as an umbrella term for everything psychological. So sometimes you will encounter references to mental processes as types of "behaviour". This is not exactly accurate, but acceptable.

Note that the definition of psychology does not specify human behaviour or mental processes. This is because research with non-human animals is also an integral part of psychology. Since humans are just a stage in the continuous process of evolution, the study of animals may inform our understanding of human behaviour (and mental processes).

IB psychology is an academic discipline with an emphasis on rigorous research and scientific knowledge, but psychology is broader than pure academics and research. When people think about psychology many imagine counsellors and psychotherapists, practitioners who work with individual clients. University workers in lab coats conducting research is not the first thing that comes to mind. However, IB psychology focuses on academic knowledge and scientific research rather than counselling skills. This is because thorough understanding of psychological concepts and being able to think critically about psychological phenomena is of paramount importance in all spheres of psychology,

including counselling. It makes perfect sense to start with building these skills, much like the need to study aerodynamics before you are allowed to pilot an airplane.

## Research methodology: quantitative and qualitative methods

All research methods used in psychology can be categorized as either quantitative or qualitative. Data in quantitative research comes in the form of numbers. The aim of quantitative research is usually to arrive at numerically expressed laws that characterize behaviour of large groups of individuals (that is, universal laws). This is much like the aim of the natural sciences in which it has been the ideal for a long time to have a set of simple rules that describe the behaviour of all material objects throughout the universe (think about laws of gravity in classic Newtonian physics, for example). In philosophy of science such orientation on deriving universal laws is called the **nomothetic approach**.

**Quantitative research** operates with **variables**. A variable (“something that can take on varying values”) is any characteristic that is objectively registered and quantified. Since psychology deals with a lot of “internal” characteristics that are not directly observable, they need to be **operationalized** first. For this reason, there’s an important distinction between **constructs** and **operationalizations**.

A construct is any theoretically defined variable, for example, violence, aggression, attraction, memory, attention, love, anxiety. To define a construct, you give it a definition which delineates it from other similar (and dissimilar) constructs. Such definitions are based on theories. As a rule constructs cannot be directly observed: they are called constructs for a reason—we have “constructed” them based on theory.

To enable research, constructs need to be operationalized. Operationalization of a construct means expressing it in terms of observable behaviour. For example, to operationalize verbal aggression you might look at “the number of insulting comments per hour” or “the number of swear words per 100 words in the most recent Facebook posts”. To operationalize anxiety you

might look at a self-report score on an anxiety questionnaire, the level of cortisol (the stress hormone) in the bloodstream or weight loss. As you can see, there are usually multiple ways in which a construct may be operationalized; the researcher needs to use creativity in designing a good operationalization that captures the essence of the construct and yet is directly observable and reliably measurable. As you will see throughout examples in this book, it is often a creative operationalization that makes research in psychology outstanding.

### ATL skills: Research and communication

In small groups think of operationalizations of the following constructs: belief in God, assertiveness, shyness, pain, love, friendship, prejudice, tolerance to uncertainty, intelligence, wisdom.

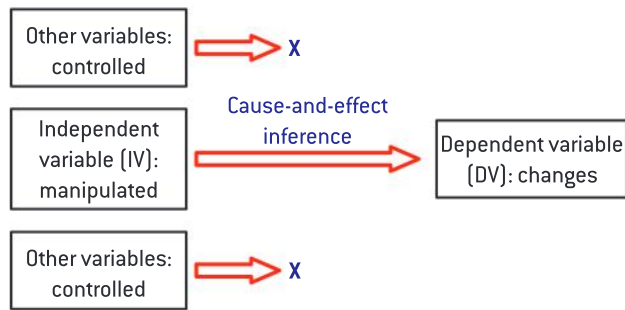
Is it equally easy to operationalize them?

Discuss each other’s operationalizations and outline their strengths and limitations.

There are three types of quantitative research.

- Experimental studies.** The experiment in its simplest form includes one **independent variable** (IV) and one **dependent variable** (DV), while the other potentially important variables are controlled. The IV is the one manipulated by the researcher. The DV is expected to change as the IV changes. For example, if you want to investigate the effect of psychotherapy on depression, you might randomly assign participants to two groups: the experimental group will receive psychotherapy while the control group will not. After a while you might measure the level of depression by conducting a standardized clinical interview (diagnosis) with each of them. In this case the IV is psychotherapy. You manipulate the IV by changing its value: yes or no. The DV is depression; it is operationalized through the standardized diagnostic procedure. If the DV is different in the two groups, you may conclude that a change in the IV “caused” a change in the DV. This is why the experiment is the only method that allows **cause-and-effect inferences**.





▲ Figure 1.2 Cause-and-effect inference

- **Correlational studies.** Correlational studies are different from experiments in that the researcher does not manipulate any variables (there are no IVs or DVs). Variables are measured and the relationship between them is quantified. For example, if you want to establish if there is any relationship between violent behaviour of adolescents and how much time they spend watching violent television shows, you may recruit a sample of adolescents and measure their violent behaviour (by self-report, by ratings from

classmates or even by observation in a natural setting) and the average number of hours per day spent watching violent television shows. Then you can correlate these two variables using a formula. Suppose you obtained a large positive correlation. This means that there's a trend in the data: the more time an adolescent spends watching violent shows, the more violent he or she is. However, you cannot make cause-and-effect inferences from correlational studies. Since you did not manipulate one of the variables, you do not know the direction of influence. It could be the case that watching violence influences violent behaviour (this would probably be the most popular, intuitive assumption). However, it is also possible that adolescents who behave violently choose to watch violent television programmes. Or there could even be a third variable (for example, low self-esteem) that influences both violent behaviour and watching violence on television. What you observe "on the surface" is just that—"co-relation", the fact that one variable changes as the other one changes.

### ATL skills: Communication and social

In small groups come up with results of fictitious studies that would demonstrate either correlation or causation. Here are two examples.

1. In a group of adults we measured their attitudes to horror films and the number of siblings they have. We found that the more siblings you have, the more you like horror films.
2. We told one group of astronauts that their mission would start in a month and the other group that the mission would start in a year. We measured anxiety and found that it was higher in the group of astronauts who expected the mission to start in a month.

As you go through your list of fictitious studies, the other groups will have to say whether the study shows correlation or causation.

- **Descriptive studies.** In descriptive studies relationships between variables are not investigated, and the variables are approached separately. An example of a descriptive quantitative study would be a public opinion survey. We ask questions (for example, "Do you support the current policies of the President?") and we are interested in the distribution of answers to this particular question. Descriptive studies are often used in sociology and they are sometimes used in psychology to conduct a broad investigation of a phenomenon before "delving deeper" into the specifics.

**Qualitative research** is different. Its main focus is an in-depth study of a particular phenomenon. "In-depth" entails going beyond what can be objectively measured and quantified into the realm of human experiences, interpretations and meanings. Qualitative research makes use of such data collection methods as interviews or observations. Data comes in the form of texts: interview transcripts, observational notes, and so on. Interpretation of data involves a degree of subjectivity, but analysis is deeper than we can usually achieve through quantitative approaches. In philosophy of science such



orientation on an in-depth analysis of a particular case or phenomenon (without trying to derive

universally applicable laws) is called the **idiographic approach**.

Parameter	Quantitative research	Qualitative research
<b>Aim</b>	Nomothetic approach: derive universally applicable laws	Idiographic approach: in-depth understanding of a particular case or phenomenon
<b>Data</b>	Numbers	Texts
<b>Focus</b>	Behavioural manifestations (operationalizations)	Human experiences, interpretations, meanings
<b>Objectivity</b>	More objective (the researcher is eliminated from the studied reality)	More subjective (the researcher is included in the studied reality)

▲ Table 1.1 Quantitative versus qualitative research

Qualitative research methods that we will discuss in this chapter are:

- observation
- interview
- focus group
- case study
- content analysis.

## Sampling, credibility, generalizability and bias in research

Sampling, credibility, generalizability and bias are some of the characteristics used to describe a research study and make a judgment of its quality. These characteristics are universal for social sciences, but they can be approached very differently by quantitative and qualitative researchers, sometimes even with distinctly different sets of terms to express the same ideas. So it is important that you understand both these overarching concepts and the way they are broken down in quantitative as compared to qualitative research. Let's start with the overarching concepts.

**A sample** is the group of individuals taking part in the research study. **Sampling** is the process of finding and recruiting individuals for the study. There are different sampling techniques, and it is important to be aware of their strengths and limitations as sampling may affect the results of the study. For example, if the aim of your research is to see if anxiety correlates with aggression in teenagers (in general), but you only sample teenagers from one school in a criminal neighbourhood, your sampling technique will have important implications

for the conclusions you will be able to make. Similarly, if you study political views of unemployed people and you recruit your sample by asking a small number of participants to bring their friends (and possibly friends of friends), you might end up with a limited sample because people of similar political views are more likely to be friends with each other.

**Credibility** refers to the degree to which the results of the study can be trusted to reflect the reality. It is closely linked to **bias**, because the results of the study do not reflect reality if there was some sort of bias in it. There are a lot of "traps" that a researcher can walk into. For example, in an interview, while the researcher believes the interviewee's responses to be true, participants may actually guess the aim of the study and respond in a way that they think the researcher is expecting them to. Or researchers themselves, being interested in confirming their hypothesis, may selectively notice supporting evidence and unintentionally ignore contradicting evidence. If there is indication that potential sources of bias were, to the best of our knowledge and abilities, controlled or eliminated, credibility of the research study is believed to be high. Quantitative and qualitative research approaches to credibility and bias are distinctly different, although they overlap in some aspects.

**Generalizability** refers to the extent to which the results of the study can be applied beyond the sample and the settings used in the study itself. Sometimes, especially in quantitative research, you want to generalize findings from the sample to a much wider group of people (called "population") because your aim is to discover universal laws of behaviour. Sometimes the research study is conducted in artificial settings

(for example, a laboratory experiment), but you want to believe that people will behave the same way in their natural setting in daily life too. In any case, generalizability is an important aspect in the interpretation of findings. Again, the ways in which quantitative and qualitative research studies approach generalizability of findings is distinctly different.

The table below gives you an overview of the main concepts used to characterize sampling, generalizability, credibility and bias in experimental, correlational and qualitative research. As you read on, you will understand these concepts better. Refer to this table from time to time so that you place them clearly in the general framework.

**Overview table:  
Sampling, generalizability, credibility and bias in qualitative and quantitative research**

Overarching concepts	Quantitative research		Qualitative research
	Experimental studies	Correlational studies	
Sampling	Random Stratified Self-selected Opportunity	Same	Quota sampling Purposive sampling Theoretical sampling Snowball sampling Convenience sampling
Generalizability	External validity: – Population validity – Ecological validity Construct validity	Population validity Construct validity	Sample-to-population generalization Case-to-case generalization Theoretical generalization
Credibility	Internal validity: to what extent is the DV influenced by the IV and not some other variable?  Controlling confounding variables: eliminating or keeping constant in all conditions	No special term used: “validity” and “credibility” can be used interchangeably  Credibility is high if no bias occurred	Credibility = trustworthiness. To what extent do the findings reflect the reality?  Triangulation Establishing a rapport Iterative questioning Reflexivity Credibility checks Thick descriptions
Bias	Threats to internal validity: – Selection – History – Maturation – Testing effect – Instrumentation – Regression to the mean – Experimental mortality – Experimenter bias – Demand characteristics	On the level of measurement of variables: depends on the method of measurement  On the level of interpretation of findings: – Curvilinear relationships – The third variable problem – Spurious correlations	Participant bias: – Acquiescence – Social desirability – Dominant respondent – Sensitivity Researcher bias: – Confirmation bias – Leading questions bias – Question order bias – Sampling bias – Biased reporting

▲ Table 1.2

# Quantitative research: the experiment

## Inquiry questions

- Why do experiments allow cause-and-effect inferences?
- How can bias in experimental research be prevented?
- How can findings from a small group of people be generalized to an entire population?
- How can experiments be designed?

## What you will learn in this section

- Confounding variables
- Sampling in the experiment
  - Representativeness
  - Random sampling
  - Stratified sampling
  - Opportunity sampling
  - Self-selected sampling
- Experimental designs
  - Independent measures design
  - Matched pairs design; matching variable
  - Repeated measures design; order effects; counterbalancing
- Credibility and generalizability in the experiment: types of validity
  - Construct validity
  - Internal validity
  - External validity: population and ecological
- Bias in experimental research: threats to internal validity
  - Selection
  - History
  - Maturation
  - Testing effect
  - Instrumentation
  - Regression to the mean
  - Mortality
  - Demand characteristics
  - Experimenter bias
- Quasi-experiments versus true experiments
- Natural experiments and field experiments

## Confounding variables

As we mentioned, the experiment is the only method that allows researchers to make cause-and-effect inferences. This is achieved by defining the independent variable (IV) and the dependent variable (DV), manipulating the IV and observing how the DV changes in response to this manipulation.

Psychological reality is very complex and the trick is to isolate the IV so that when you manipulate it, nothing else changes. Imagine, for example, that you manipulate X and observe the resulting changes in Y. However, every time you manipulate

X, you also unintentionally change Z. In reality it is Z that causes a change in Y, but you incorrectly conclude that X (your IV) is the cause of Y, thus incorrectly confirming your hypothesis. If this sounds too abstract, think about the following example: X is sleep deprivation (which you manipulate by waking up one group of participants every 15 minutes when they sleep, while the control group sleeps normally) and Y is memory performance (which you measure by a simple memory test in the morning). Without realizing that this might be an important factor, you let the control group sleep at home while the experimental group sleeps in a laboratory being

supervised by an experimenter. So there's another variable, variable Z: stress caused by the unfamiliar environment. It could be the case that in this experiment it was the unfamiliar environment (Z) that caused a reduction in memory performance (Y), rather than sleep deprivation (X).

Variables that can potentially distort the relationship between the IV and the DV (like Z in the example above) are called **confounding variables**. They contribute to bias. These variables need to be controlled, either by eliminating them or keeping them constant in all groups of participants so that they do not affect the comparison.

### Discussion

How could the researchers have controlled the confounding variable in this example?

### Exercise

Imagine you are investigating the influence of praise on the school performance of teenagers. For this experiment you need to have a sample of participants that you would split into two groups (experimental and control). In the experimental group the teacher is instructed to praise every student three times a week while in the control group the teacher is told to only praise the students once every week. At the end of the research period performance grades in the two groups are compared.

Suppose that the participants in this experiment are high school students from one of the schools in your city. Will you be able to generalize the findings to the target population, that is, teenagers in general? This depends on how representative your sample is. For this you need to take into account your target population and the aim of the research.

- The aim of the research links to the **participant characteristics** that are essential. Whatever can theoretically influence the relationship between the IV and the DV is essential. For example, cultural background may be essential for how a teenager reacts to praise (depending on that teenager's cultural attitudes to adults, teachers and authority in general). Socio-economic background may be important as well: theoretically there may be a connection between the socio-economic status of a teenager's family and their value of education. The type of school is another potentially important factor: in top schools where students pursue quality education and prestigious college placements teachers' praise may be a point of pride, whereas in public schools in criminal neighbourhoods it may lead to bullying from classmates.
- If the sample is representative, it must reflect the essential characteristics of the target population. Is the sample of teenagers from one school in our example sufficient to reflect all these characteristics? No, because it does not represent the variation of cultural backgrounds, socio-economic backgrounds and types of schools found in the population.
- If the sample is not representative of the essential characteristics of the target population, there are two ways to fix it: either keep sampling or narrow down the target population and do not claim that the research findings are more generalizable than they really are.

Given the aim of the study, how would you increase representativeness of your sample?

## Sampling in the experiment

Being a truly nomothetic method, the experiment aims at discovering universal laws of behaviour applicable to large groups of people across a variety of situations. This makes relevant the distinction between the **sample** and the **target population**. The target population is the group of people to which the findings of the study are expected to be generalized. The sample is the group of people taking part in the experiment itself. How can we ensure that whatever results are obtained in the sample can be generalized to the target population? We do this through **representativeness**—the key property of a sample. A sample is said to be representative of the target population if it reflects all its essential characteristics.





There is no quantitative way to establish representativeness of a sample and it is always the expert decision of a researcher to say whether a particular characteristic is essential or not. This is done on the basis of prior knowledge from published theories and research studies. In any case the choice of the target population needs to be well justified and explicitly explained.

Several **sampling techniques** can be used in an experiment. The choice depends on the aim of the research, available resources and the nature of the target population.

- Random sampling.** This is the ideal approach to make the sample representative. In random sampling every member of the target population has an equal chance of becoming part of the sample. With a sufficient sample size this means that you take into account all possible essential characteristics of the target population, even the ones you never suspected to play a role. Arguably, a random sample of sufficient size is a good representation of a population, making the results easily generalizable. However, random sampling is not always possible for practical reasons. If your target population is large, for example, all teenagers in the world, it is impossible to ensure that each member of this population gets an equal chance to enter your sample. Being based in Europe, you cannot just create a list of all teenagers in the world, randomly select a sample and then call Lynn from Fiji to come and join your experiment. In

this case you either believe that cross-cultural differences are not essential (for your hypothesis) or narrow down your target population. On the other hand, if your target population is students from your school, it is perfectly possible to create the full list of students and select your participants randomly from this list. An example of random sampling strategy is a pre-election telephone survey where participants are selected randomly from the telephone book (or a random selection of Facebook profiles). Even in this case, though, you have to admit that the target population is not all the citizens of a particular country; it is all the citizens of the country who own a telephone (or have a Facebook profile).

- Stratified sampling.** This approach is more theory-driven. First you decide the essential characteristics the sample has to reflect. Then you study the distribution of these characteristics in the target population (for this you may use statistical data available from various agencies). Then you recruit your participants in a way that keeps the same proportions in the sample as is observed in the population. For example, imagine that your target population is all the students in your school. The characteristics you decide are important for the aim of the study are age (primary school, middle school, high school) and grade point average—GPA (low, average, high). You study school records and find out the distribution of students across these categories:

	Low GPA	Average GPA	High GPA	Total
Primary school	0%	10%	10%	20%
Middle school	5%	30%	15%	50%
High school	5%	20%	5%	30%
Total	10%	60%	30%	100%

▲ Table 1.3

For a stratified sample you need to ensure that your sample has the same proportions. For every cell of this table you can either sample randomly or use other approaches (see below). In any case, what makes stratified sampling special is that it is theory-driven and it ensures that theory-defined essential characteristics of the population are fairly and equally represented in the sample. This may be the ideal choice when you are certain about essential participant characteristics and when available sample sizes are not large.

- Convenience (opportunity) sampling.** For this technique you recruit participants that are more easily available. For example, university students are a very popular choice because researchers are usually also university professors so it is easy for them to find samples there. Jokingly, psychology has been sometimes referred to as a study of “US college freshmen and white rats”. There could be several reasons for choosing convenience sampling. First, it is the technique of choice when financial resources and time are limited. Second, there

could be reasons to believe that people are not that different in terms of the phenomenon under study. For example, if you study the influence of caffeine on attention, there are reasons to believe that results will be similar cross-culturally, and it might be a waste of time to use a stratified or a random sample. Finally, convenience sampling is useful when wide generalization of findings is not the primary goal of your research, for example, if you are conducting an exploratory study and you are not sure the hypothesis will be supported by evidence. If the hypothesis will not “work” in a small sample, why waste time testing it in a representative sample? Or you are replicating someone else’s research and your aim is to see if the universal law (that was discovered by this someone) will hold true in your specific sample, thus trying to falsify prior theory. The limitation of convenience sampling is, of course, lack of representativeness.

- **Self-selected sampling.** This refers to recruiting volunteers. An example of this approach is advertising the experiment in a newspaper and using the participants who respond to the advert. The strength of self-selected sampling is that it is a quick and relatively easy way to recruit individuals while at the same time having wide coverage (many different people read newspapers). The most essential limitation, again, is representativeness. People who volunteer to take part in experiments may be more motivated than the general population, or they may be looking for the incentives (in many studies participants are financially rewarded for their time).

### Exercise

Now that you know what sampling strategies can be used in an experiment, how would you change your approach to recruiting a sample for the investigation of the influence of praise on school performance of teenagers?

## Experimental designs

Experiments always involve manipulating some variables and measuring the change in others. But the specific ways in which this can be organized

differ depending on the aims of the research. The organization of groups and conditions in an experiment is known as the experimental design, and there are three basic types of experimental design.

**Independent measures design** involves random allocation of participants into groups and a comparison between these groups. In its simplest form, you randomly allocate participants from your sample into the experimental group and the control group. Then you manipulate the experimental conditions so that they are the same in the two groups except for the independent variable. After the manipulation you compare the dependent variable in the two groups.

### ATL skills: Research

Consider 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.

The rationale behind random group allocation is that all potential confounding variables cancel each other out. If the groups are not equivalent at the start of the experiment, you will be comparing apples to oranges. Imagine that you are testing the hypothesis that praise at school improves students’ performance and for this you take two existing groups of students, with one being rarely praised by their teachers and the other one often praised. Arguably, the groups might not be equivalent: they have different experiences with the teachers, different ingroup values and habits, and so on—but to account for all these potentially important factors is impossible.

Conversely, when the group sizes are sufficiently large and allocation is completely random, chances are that groups will be equivalent—the larger the sample, the higher the chance.

Of course, there could be more than two groups, depending on how many IVs you use and how many levels each variable has. In the above example, you could use more than one IV: the influence of praise and the allocation of homework on school performance. With two levels for each of these IVs you would need to randomly allocate participants into four groups:

	Homework given	Homework not given
Rarely praised	1	2
Frequently praised	3	4

▲ Table 1.4

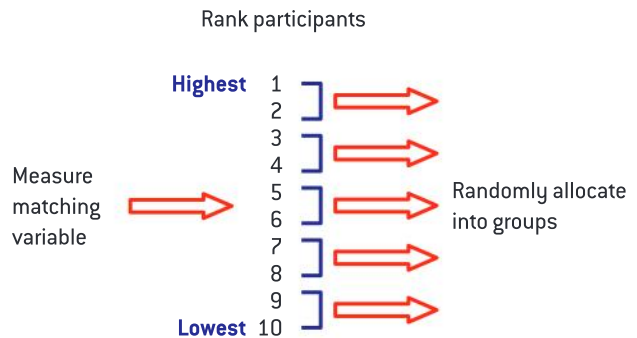
This experimental design with two IVs, each with two levels, is quite frequently used in psychological experiments. It is known as a 2 × 2 experimental design. Of course you can think of other combinations: 2 × 3 (two IVs, three levels in each), 3 × 2 (three IVs, two levels in each), 4 × 4 (four IVs, four levels in each). The more cells you have in this table, the larger the sample you need, so at some point it becomes impractical to increase the number of groups.

To summarize, regardless of the number of IVs and their levels, an experiment follows an independent measures design when the IV is manipulated by randomly allocating participants into groups. This allows us to assume that the groups are equivalent from the start so whatever difference we observe at the end of the experiment must have been caused by our experimental manipulation.

**Matched pairs design** is similar to independent measures. The only difference is that instead of completely random allocation, researchers use matching to form the groups.

To illustrate matching, let’s consider an example. Suppose you are conducting a study of the effect of sleep deprivation on memory. For this you need two groups of participants. One of the groups will sleep peacefully in the laboratory and the other group will be woken up every 15 minutes. In the morning you will give both groups a memory test and compare their performance. You suspect that there is one confounding variable that may influence the results: memory abilities. Some people generally have better memory than others, therefore it is important to you that the two groups at the start of the experiment are equivalent in their memory abilities. Random allocation will usually make that happen, but you only have 20 participants (10 in each group). With a small sample like this there is a chance that random allocation will not work. So you want to control the equivalence of memory abilities “manually” while leaving everything else to random chance. For this you test memory abilities in your participants prior to the experiment. Then you rank participants

according to their memory abilities (for example, from the highest to the lowest). Then you take the first two participants from the top of the list and randomly allocate one of them to the experimental group and the other one to the control group. You take the next two participants and repeat the procedure for the rest of the list. The two resulting groups are certainly equivalent in terms of memory abilities and probably (due to random chance) equivalent in all other characteristics.



▲ Figure 1.3 Matched pairs design

The variable that is controlled (memory abilities in the example above) is called the **matching variable**. Matched pairs designs are preferred when:

- the researcher finds it particularly important that the groups are equivalent in a specific variable
- the sample size is not large, therefore there is a chance that random allocation into groups will not be sufficient to ensure group equivalence.

**Repeated measures design** is used when the goal is to compare conditions rather than groups of participants. The same group of participants is exposed to two (or more) conditions, and the conditions are compared. For example, imagine your aim is to investigate the effect of classical music on learning. You ask your participants to learn a list of trigrams (meaningless combinations of three letters such as HPX, LJW) for 10 minutes in silence and register the number of trigrams correctly recalled. Then you ask the same participants to learn a different list of trigrams for another 10 minutes, but this time with classical music playing in the background. You compare results from the first and the second trial.

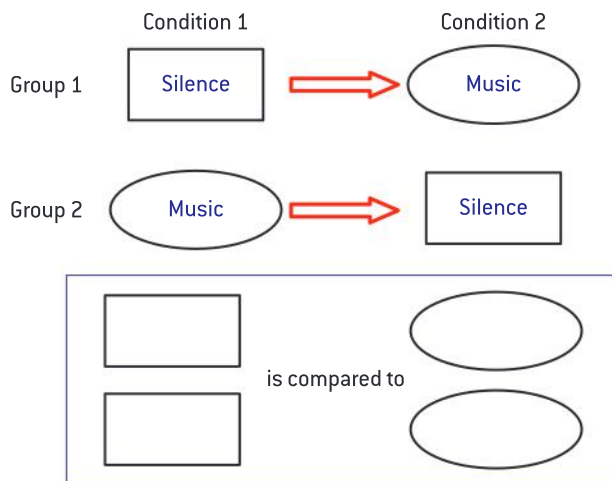
The problem with repeated measures designs is that they are vulnerable to **order effects**: results



may be different depending on which condition comes first (for example, silence then classical music or classical music then silence). Order effects may appear due to various reasons, such as the following.

- **Practise:** participants practise, improve their on-task concentration and become more comfortable with the experimental task during the first trial. Their performance in the second trial increases.
- **Fatigue:** participants get tired during the first trial, and their concentration decreases. Their performance in the second trial decreases.

To overcome order effects researchers use **counterbalancing**. Counterbalancing involves using other groups of participants where the order of the conditions is reversed. For our example, two groups could be used: one given the sequence “silence then music” and one given the sequence “music then silence”. It is important to note that comparison will still be made between conditions, not between groups. Data from group 1 condition 1 will be collated with data from group 2 condition 2, and vice versa. These two collated data sets will be compared.



▲ Figure 1.4 Counterbalancing

An advantage of repeated measures designs is that people are essentially compared to themselves, which overcomes the influence of **participant variability** (differences between the groups before the experiment starts). It makes the comparison more reliable. Another advantage following from this is that smaller sample sizes are required.

## Credibility and generalizability in the experiment: types of validity

As you have seen, credibility and generalizability are overarching terms that are used to characterize the quality of research studies. When it comes to experiments specifically, these terms are very rarely used. Instead the quality of experiments is characterized by their construct, internal and external validity.

**Construct validity** characterizes the quality of operationalizations. As you know, the phenomenon under study is first defined theoretically as a construct and then expressed in terms of observable behaviour (operationalization). Operationalization makes empirical research possible. At the same time when results are interpreted research findings are linked back to constructs. Moving from an operationalization to a construct is always a bit of a leap. Construct validity of an experiment is high if this leap is justified and if the operationalization provides sufficient coverage of the construct. For example, in some research studies anxiety was measured by a fidgetometer, a specially constructed chair that registers movements at various points and so calculates the amount of “fidgeting”. Subjects would be invited to the laboratory and asked to wait in a chair, not suspecting that the experiment has already started. The rationale is that the more anxious you are, the more you fidget in the chair. Are the readings of a fidgetometer a good operationalization of anxiety? On the one hand, it is an objective measure. On the other hand, fidgeting may be a symptom of something other than anxiety. Also the relationship between anxiety and increased fidgeting first has to be demonstrated in empirical research.

**Internal validity** characterizes the methodological quality of the experiment. 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. In other words, internal validity links directly to bias: the less bias, the higher the internal validity of the experiment. Biases in the experiment (threats to internal validity) will be discussed below.

**External validity** characterizes generalizability of findings in the experiment. There are two types of external validity: population validity and ecological validity. **Population validity** refers to the extent

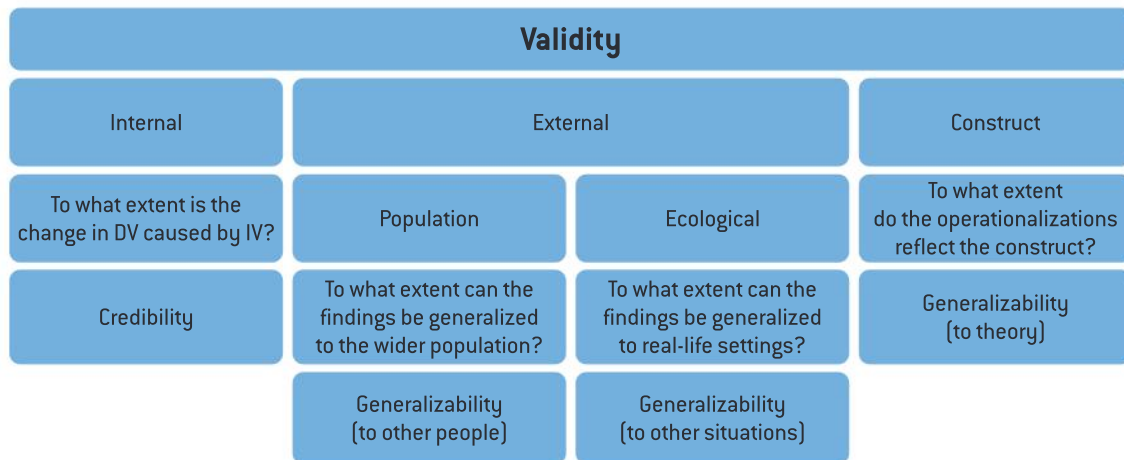




to which findings can be generalized from the sample to the target population. Population validity is high when the sample is representative of the target population and an appropriate sampling technique is used. **Ecological validity** refers to the extent to which findings can be generalized from the experiment to other settings or situations. It links to the artificiality of experimental conditions. In highly controlled laboratory experiments subjects often find themselves in situations that do not resemble their daily life. For example, in memory experiments they are often asked to memorize long lists of trigrams. To what

extent can findings from such studies be applied to everyday learning situations?

There is an inverse relationship between internal validity and ecological validity. To avoid bias and control for confounding variables, you make the experimental procedures more standardized and artificial. This reduces ecological validity. Conversely, in an attempt to increase ecological validity you may allow more freedom in how people behave and what settings they choose, but this would mean that you are losing control over some potentially confounding variables.



▲ Figure 1.5 Validity of experiments

**Exercise**

- Leaf through this book (consider the units on the biological, cognitive or sociocultural approach to behaviour), find a description of any experimental study and analyse its construct, internal and external validity. If you feel that you do not have enough detail, you could find more information on the study online, or even read the original article.
- Present the results of your analysis in class.

**Bias in experimental research: threats to internal validity**

Bias in experimental research comes in the form of confounding factors that may influence the cause-and-effect relationship between the IV and the DV, decreasing internal validity. Below you will find a description of several common sources of threat to internal validity, based on Campbell (1969).



▲ Figure 1.6 Sources of threat to internal validity

1. **Selection.** This occurs if for some reason groups are not equivalent at the start of the experiment: apart from the planned IV-related difference, they differ in some other variable. As a result, we cannot be sure if the post-experiment differences between groups reflect the influence of the IV or this other variable. Selection occurs in independent measures and matched pairs designs in case group allocation was not completely random.
2. **History.** This refers to outside events that happen to participants in the course of the experiment. These outside events become a problem when they can potentially influence the DV or are not evenly distributed in the comparison groups. History is especially important in lengthy experiments where the DV is measured sometime after the onset of the study. For an example of history-related bias think of a memory experiment where participants are required to memorize long lists of words and the experiment is conducted in two groups (experimental and control) simultaneously in two different rooms on the opposite sides of a school. As the experiment begins, there is some noise coming from road construction outside. The control group is closer to the construction site so the noise in their room is louder. Since distracting noise can affect memory performance and levels of noise were not equal in the two groups, resulting differences in the DV may reflect the influence of the IV as well as the confounding variable (noise). To counteract history as a threat to internal validity such confounding variables should be either eliminated or kept constant in all comparison groups (for example, change the rooms so that they are both on the same side of the school building).
3. **Maturation.** In the course of the experiment participants go through natural developmental processes, such as fatigue or simply growth. For example, suppose you are piloting a psychological training programme to increase assertiveness in middle school students. You measure assertiveness at the start, conduct the training programme for several months and measure assertiveness again. The resulting increase of assertiveness may be due to either the IV (the training) or simply to the fact that the middle school students grew up a

little and naturally became more assertive. The counteracting strategy would be using a control group (the same time period, the same measurements but no training sessions).

4. **Testing effect.** The first measurement of the DV may affect the second (and subsequent) measurements. For example, suppose you are investigating the effectiveness of a video to reduce test anxiety in primary school children. For this your participants take an ability test preceded by a self-report anxiety measure at time A. They then watch your specially designed video and repeat the procedure (test and self-report anxiety measure) at time B. The difference in anxiety between time A and time B may be the result of both the video and the fact that it is their second time taking the test—they are more familiar with the format and therefore may be naturally less anxious. A solution to this is to use a control group where you show a neutral video of the same duration. Suppose you get the following results:

Group	Test anxiety (on a scale 0–100)	
	Before Test 1	Before Test 2
Experimental (specially designed video)	90	55
Control (neutral video)	90	70

▲ Table 1.5

Analysis of these results can reveal that a reduction of anxiety by 20 points is probably due to the testing effect; however, over and above that there is a 15-point anxiety effect of the specially designed video.

In repeated measures designs testing effect is a special case of order effects, and counterbalancing is used to control for it.

5. **Instrumentation.** This effect occurs when the instrument measuring the DV changes slightly between measurements. For psychology this becomes relevant when you consider that an “instrument of measurement” is often a human observer. Suppose you are investigating bullying on a school campus during breaks. You are looking at two groups of students who are exposed to different experimental conditions. If

you observe group 1 in the morning and group 2 in the afternoon, you might be more tired in the afternoon and miss some important behaviours. If you observe one of the groups during a short break and the other one during the lunch break, observations during the lunch break may be less accurate because it is more crowded. To avoid this researchers should try to standardize measurement conditions as much as possible across all comparison groups and all observers.

6. **Regression to the mean.** This is an interesting source of bias that becomes a concern when the initial score on the DV is extreme (either low or high). Extreme scores have a purely statistical tendency to become more average on subsequent trials. Suppose you have designed anxiety reduction training for students. To test its effectiveness, you administer an anxiety questionnaire in a group of students and select a sample of students who have the largest score (for example, 80–100 on a 100-point scale). With these students you then conduct your training session and measure their anxiety again. Even if we assume that testing effects are not an issue, we would expect extremely anxious students to naturally become less anxious even without the training session. To put it more precisely, the probability that extremely anxious students will become even more anxious is less than the probability that they will become less anxious. This means that statistically a reduction of anxiety should be expected. A counter-measure is a control group with the same starting average anxiety level and measurements at the same point of time, but without the intervention.
7. **Experimental mortality.** This refers to the fact that some participants drop out during an experiment, which may become a problem if dropouts are not random. Suppose you are investigating the influence of emotion on ethical decision-making. For this you give your participants a number of scenarios of the type “Would you kill 1 person to save 1000?” In the control group the description of this “one person” is neutral, but in the experimental group this is someone they know personally, so there is more emotional involvement. You hypothesize that people will be less likely to be utilitarian in their decision-making when they are personally involved (note that this research would create distress among participants and so raises ethical issues; it is quite possible such a study would not be approved by the ethics committee). Suppose that several participants in the experimental group refuse to continue participation and drop out, more so than in the control group. Ethical issues aside, this presents a methodological issue as well: even if the two groups were equivalent at the start of the experiment, they may be non-equivalent now. There appears a confounding variable (sensitivity) which is disproportionately represented in the two groups. There is no reliable way to counteract experimental mortality other than designing experimental conditions in such a way that participants would not feel the need to drop out.
8. **Demand characteristics.** This refers to a situation in which participants understand the purpose of the experiment and change their behaviour subconsciously to fit that interpretation. In other words, they behave in ways that they think the experimenter expects. This can happen for various reasons, for example, participants may feel that they will somehow be evaluated and so behave in a socially desirable way. To avoid demand characteristics, deception may be used to conceal the true purpose of the study (however, deception raises ethical issues—see below). You can consider using post-experimental questionnaires to find out to what extent demand characteristics may have influenced the results (this strategy does not prevent demand characteristics but just estimates their impact). Note that in repeated measures designs demand characteristics are a larger threat because participants take part in more than one condition and so have greater opportunities to figure out or guess the aim of the study.
9. **Experimenter bias.** This refers to situations in which the researcher unintentionally exerts an influence on the results of the study, for example, the Clever Hans case discussed above. Existence of this bias was first rigorously supported by Rosenthal and Fode (1963). In this experiment rats were studied for their maze-running performance. Rats were split into two groups at random, but the laboratory assistants (psychology students) were told that one of the groups was “maze-bright” and

the other one was “maze-dull” and that this difference in ability was genetic. Laboratory assistants had to follow a rigorous and standardized experimental procedure in which rats were tested on their performance in learning the maze task. This was supposed to be an identical study conducted with identical rats, but results showed that the rats labelled “maze-dull” performed significantly worse than the ones labelled “maze-bright”. It was concluded that the result was an artifact: it was caused by experimenter bias rather than any genuine differences between the groups of rats. Post-experiment investigations revealed that experimenter bias was not intentional or conscious. The results were induced by subtle differences in the way laboratory assistants handled the rats. For example, without realizing it, assistants handled “maze-bright” rats for slightly longer and so stress was more reduced for these rats than for “maze-dull” rats. A counter-measure against experimenter bias

is using so-called **double-blind designs** where information that could introduce bias is withheld both from the participants and from the people conducting the experiment. The study of Rosenthal and Fode would have been double-blind if the laboratory assistants had not been told which group of rats had which label.

### Exercise

Once again leaf through this book and find a description of any experimental study.

- To what extent was this experimental study susceptible to one of the sources of threat to internal validity? What does it tell you about credibility of the study?
- If you do not have enough detail, find more information on the study online, or even read the original article.
- Present the results of your analysis in class.

### ATL skills: Self-management

Athabasca University has a great learning resource on threats to internal validity. One tutorial consists of two parts, where part 1 is the theoretical background and definitions and part 2 is a practical exercise involving the analysis of 36 hypothetical experiments.

If you want to practise identifying potential sources of bias in experiments, you can access the tutorial here: <https://psych.athabascau.ca/open/validity/index.php>



## Quasi-experiments versus true experiments

**Quasi-experiments** are different from “true” experiments in that the allocation into groups is not done randomly. Instead some pre-existing inter-group difference is used. “Quasi” is a prefix meaning “almost”. The major limitation of a quasi-experimental design is that cause-and-effect inferences cannot be made. This is because we cannot be sure of the equivalence of comparison groups at the start of the study: pre-existing differences in one variable may be accompanied by a difference in unexpected confounding variables.

Suppose your hypothesis is that anxiety influences test performance. You have an opportunity sample of high school students. An intuitively obvious way to test this hypothesis would be to administer an anxiety questionnaire, divide the sample into two groups (anxious and non-anxious) based on

the results, and then model a testing situation and compare test performance. The IV in this study is anxiety (it has two levels) and the DV is test performance. However, the researcher does not really manipulate the IV in this study. Pre-existing differences in anxiety are used, so we cannot be sure that anxiety is the only variable that differs in the two groups. For example, it is possible that high school students with high levels of anxiety also tend to have unstable attention, and it is actually attention that influences test performance. The bottom line is that we will be able to conclude that “anxiety is linked to test performance”, but strictly speaking we will not be able to say “anxiety influences test performance”.

To test the “influence” hypothesis a true experiment would be required, so we would have to manipulate the IV. How can you manipulate anxiety? One example is splitting participants randomly into two groups and telling one of the





groups that they should expect results of their college applications later today. Anticipation of these results would probably increase anxiety in the experimental group. Then the test can be given. (Note that such an experiment would have ethical issues since it involves major deception and creates distress among participants.)

Other examples of pre-existing differences are age, gender, cultural background and occupation. Formation of experimental groups based on these variables implies a quasi-experiment. Sometimes a “true” experiment cannot be conducted because it is impossible to manipulate the IV (for example, how do you manipulate age or gender?) so quasi-experiments are justified.

In the way they are designed (superficially) quasi-experiments resemble “true” experiments, but in terms of the possible inferences (essentially) they are more like correlational studies.

### Field experiments and natural experiments

**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. The strength of field experiments is higher ecological validity as compared to experiments in a laboratory. The limitation is less control over potentially confounding variables so there is lower internal validity. An example of a field experiment is **Piliavin, Rodin and Piliavin’s** (1969) subway study in which the researchers pretended to collapse on a subway train and observed if other passengers would come to help. To manipulate the IV, some researchers were carrying a cane (the cane condition) while others were carrying a bottle (the drunk condition).

**Natural experiments**, just like field experiments, are conducted in participants’ natural environment, but here the researcher has no control over the IV—the IV occurred naturally. Ecological validity in natural experiments is an advantage and internal validity is a disadvantage owing to there being less control over confounding variables. Another advantage of natural experiments is that they can be used when it is unethical to manipulate the IV, for example, comparing rates of development in orphans that were adopted and in those who stayed in the orphanage. Since researchers do not manipulate the IV, all natural experiments are quasi-experiments.

Type of experiment	Independent variable	Settings	Can we infer causation?
True laboratory experiment	Manipulated by the researcher	Laboratory	Yes
True field experiment	Manipulated by the researcher	Real-life	Yes (but there may be confounding variables)
Natural experiment	Manipulated by the nature	Real-life	No
Quasi-experiment	Not manipulated; pre-existing difference	Laboratory or real-life	No

▲ Table 1.6

#### Exercise

Go online and find examples of quasi-experiments, natural experiments and field experiments in psychology.

# Quantitative research: correlational studies

## Inquiry questions

- What does it mean for two variables to correlate with each other?
- What should be avoided when interpreting correlations?
- Can two correlating variables be unrelated in fact?
- Can correlations show curvilinear relationships?

## What you will learn in this section

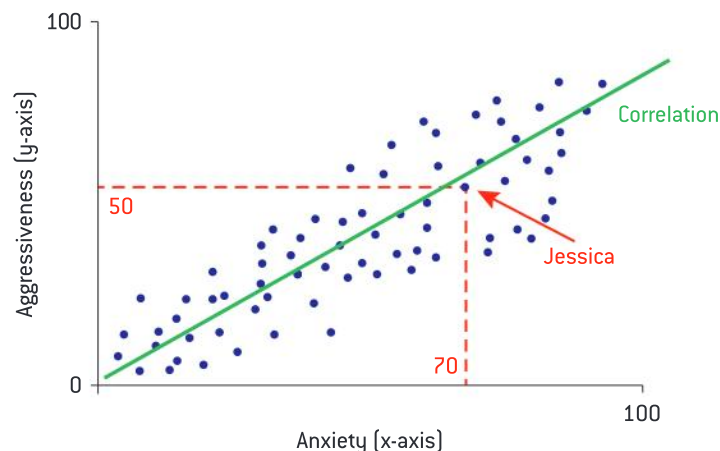
- What is a correlation?
  - Effect size
  - Statistical significance
- Limitations of correlational studies
  - Causation cannot be inferred
  - The third variable problem
  - Curvilinear relationships
  - Spurious correlations
- Sampling and generalizability in correlational studies
- Credibility and bias in correlational studies

## What is a correlation?

Correlational studies are different from experiments in that no variable is manipulated by the researcher, so causation cannot be inferred. Two or more variables are measured and the relationship between them is mathematically quantified.

The way it is done can be illustrated graphically through scatter plots. Suppose

you are interested in investigating if there is a relationship between anxiety and aggressiveness in a group of students. For this you recruit a sample of students and measure anxiety with a self-report questionnaire and aggressiveness through observation during breaks. You get two scores for each participant: anxiety and aggressiveness. Suppose both scores can take values from 0 to 100. The whole sample can be graphically represented with a scatter plot.



▲ Figure 1.7 Scatter plot

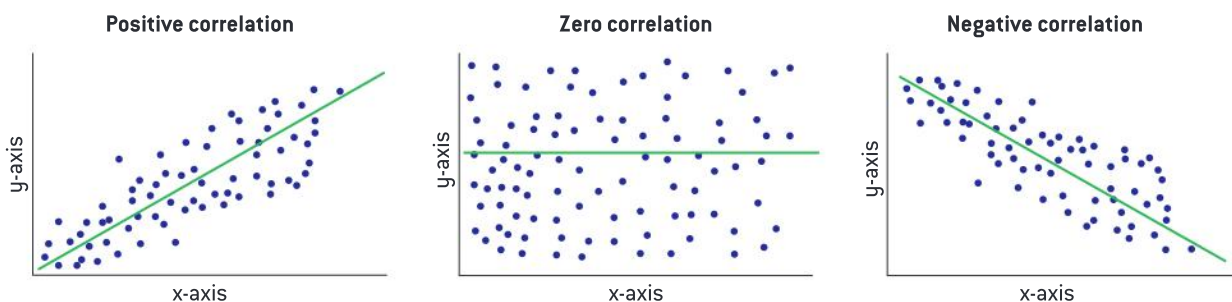


Each dot on the scatter plot represents one person. The coordinates of each dot give you the scores obtained for each of the variables. For example, Jessica’s score on anxiety is 70 (the x-axis coordinate) and her score on aggressiveness is 50 (the y-axis coordinate). The whole scatter plot looks like a “cloud” of participants in the two-dimensional space of the two variables. A **correlation** is a measure of linear relationship between two variables. Graphically a correlation is a straight line that best approximates this “cloud” in the scatter plot.

In the example above, the correlation is positive because the cloud of participants is oblong and

there is a tendency: as X increases, Y increases, so if an individual got a high score on variable X, that person probably also got a high score on variable Y, and vice versa. This is where the name “correlation” comes from: the two variables “co-relate”. Remember that correlation does not imply causation: we cannot say that X influences Y, nor can we say that Y influences X. All we know is that there is a link between them.

A correlation coefficient can vary from  $-1$  to  $+1$ . The scatter plots below demonstrate some examples:



▲ Figure 1.8 Examples of correlations

A positive correlation demonstrates the tendency for one variable to increase as the other variable increases. A negative correlation demonstrates the inverse tendency: when one variable increases the other variable decreases. The steeper the line, the stronger the relationship. A perfect correlation of 1 (or  $-1$ ) is a straight line with the slope of 45 degrees: as one variable increases by one unit, the other variable increases (or decreases) by exactly one unit. A correlation close to zero is a flat line. It shows that there is no relationship between the two variables: the fact that a person scored high or low on variable X tells us nothing about his or her score on variable Y. Graphically such scatter plots are more like a circle or a rectangle.

### Effect size and statistical significance

The absolute value of the correlation coefficient (the number from  $-1$  to  $1$ ) is called the **effect size**. How do you know if a correlation is small or large? There are widely accepted guidelines based on **Cohen’s** (1988) suggestions to interpret the effect size of correlations in social sciences.

Correlation coefficient effect size (r)	Interpretation
Less than 0.10	Negligible
0.10–0.29	Small
0.30–0.49	Medium
0.50 and larger	Large

▲ Table 1.7 Effect sizes for correlation coefficients

The effect size is not the only parameter that is important when interpreting a correlation coefficient. Another is the level of **statistical significance**. Statistical significance shows the likelihood that a correlation of this size has been obtained by chance. In other words, what is the probability that you will replicate the study with a different sample and the correlation will turn to zero? It depends on the sample size: with small samples you cannot be sure that an obtained correlation, even if it is relatively large, has not been obtained due to random chance. With large samples correlation estimates are more reliable and you can be more confident that the correlation is not a product of random chance but a genuine reflection of a relationship between the

two variables in the population. The probability that a correlation has been obtained due to random chance can be estimated. Again, there

are conventional cut-off points when results are considered to be “statistically significant” or not.

The probability that the result is due to random chance	Notation	Interpretation
More than 5%	$p = n.s.$	Result is non-significant
Less than 5%	$p < .05$	Result is statistically significant (reliably different from zero)
Less than 1%	$p < .01$	Result is very significant
Less than 0.1%	$p < .001$	Result is highly significant

▲ Table 1.8

The conventional cut-off point for statistical significance is 5%. Whatever result you obtained, if the probability that this result is pure chance occurrence is less than 5%, we assume that the

result is statistically significant, reliably different from zero and so would be replicated in at least 95 out of 100 independent samples drawn from the same target population.

### TOK

As you see, the nature of knowledge in psychology, just like the other social sciences, is probabilistic. We only know something with a degree of certainty and there is a possibility this knowledge is a product of chance.

How does that compare to the nature of knowledge in other areas such as natural sciences (physics, chemistry, biology), ethics or indigenous knowledge systems?

What can we do to increase the degree of certainty in social sciences (for example, think about replication of studies)?

When interpreting correlations one needs to take into account both the effect size and the level of statistical significance. If a correlation is statistically significant, it does not mean that it is large, because in large samples even small correlations can be significant (reliably different from zero). So, scientists are looking for statistically significant correlations with large effect sizes.

### ATL skills: Research

Correlations are denoted by the letter  $r$ . Below are some examples of results of fictitious correlational studies. See if you can interpret them using your knowledge of Cohen's effect size guidelines and levels of statistical significance:

$$r = 0.14, p = n.s.$$

$$r = 0.10, p < .05$$

$$r = 0.34, p < .01$$

$$r = 0.61, p < .001$$

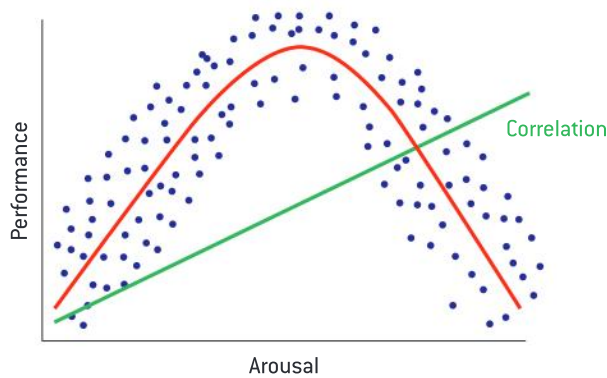
### Limitations of correlational studies

Correlational studies have several major limitations.

- As already mentioned, correlations **cannot be interpreted in terms of causation**.
- “**The third variable problem**”. There is always a possibility that a third variable exists that correlates both with X and Y and explains the correlation between them. For example, cities with a larger number of spa salons also tend to have more criminals. Is there a correlation between the number of criminals and the number of spa salons? Yes, but once you take into account the third variable, the size of the city, this correlation becomes meaningless.
- **Curvilinear relationships**. Sometimes variables are linked non-linearly. For example, a famous Yerkes-Dodson law in industrial psychology states that there is a relationship between arousal and performance: performance increases as arousal increases, but only up to a point. When levels of arousal surpass that point, performance begins to decrease.



Optimal performance is observed when levels of arousal are average. This can be seen in the scatter plot below.



▲ Figure 1.9 Arousal and performance

However, this relationship can only be captured by looking at the graph. Since correlation coefficients are linear, the best they could do is to find a straight line that fits best to the scatter plot. So, if we were using correlational methods to find a relationship between arousal and performance, we would probably end up obtaining a small to medium correlation coefficient. Psychological reality is complex and there are a lot of potentially curvilinear relationships between variables, but correlational methods reduce these relationships to linear, easily quantifiable patterns.

- **Spurious correlations.** When a research study involves calculating multiple correlations between multiple variables, there is a possibility that some of the statistically significant correlations would be the result of random chance. Remember that a statistically significant correlation is the one that is different from zero with the probability of 95%. There is still a 5% chance that the correlation is an artifact and the relationship actually does not exist in reality. When we calculate 100 correlations and only pick the ones that turned out to be significant, this increases the chance that we have picked spurious correlations.

## Sampling and generalizability in correlational studies

Sampling strategies in correlational research are the same as in experiments. First the target population is identified depending on the aims of the study and then a sample is drawn from the

population using random, stratified, opportunity or self-selected sampling.

Generalizability of findings in correlational research is directly linked with sampling and depends on representativeness of the sample. Again, this is much like population validity in experiments.

## Credibility and bias in correlational research

Bias in correlational research can occur on the level of variable measurement and on the level of interpretation of findings.

On the level of measurement of variables, various biases may occur and they are not specific to correlational research. For example, if observation is used to measure one of the variables, the researcher needs to be aware of all the biases inherent in observation. If questionnaires are used to measure variables, biases inherent in questionnaires become an issue. The list goes on.

On the level of interpretation of findings, the following considerations represent potential sources of bias.

- Curvilinear relationships between variables (see above). If this is suspected, researchers should generate and study scatter plots.
- “The third variable problem”. Correlational research is more credible if the researcher considers potential “third variables” in advance and includes them in the research in order to explicitly study the links between X and Y and this third variable.
- Spurious correlations. To increase credibility, results of multiple comparisons should be interpreted with caution. Effect sizes need to be considered together with the level of statistical significance.

### ATL skills: Self-management

Go back to the overview table [Table 1.2]. Compare and contrast sampling, generalizability, credibility and bias in correlational research with those in experimental research.

- In what aspects are the approaches different?
- In what aspects are they the same?
- Are there any aspects where the ideas are similar but the terminology differs?

# Qualitative research

## Inquiry questions

- To what extent can findings from qualitative research be generalized?
- How can credibility of qualitative research studies be ensured?
- What are the differences and similarities in how qualitative and quantitative research approaches sampling, credibility, generalizability and bias?

## What you will learn in this section

- Credibility in qualitative research
  - Triangulation: method, data, researcher, theory
  - Rapport
  - Iterative questioning
  - Reflexivity: personal, epistemological
  - Credibility checks
  - Thick descriptions
- Bias in qualitative research
- Participant bias
  - Acquiescence bias
  - Social desirability bias
  - Dominant respondent bias
  - Sensitivity bias
- Researcher bias
  - Confirmation bias
  - Leading questions bias
  - Question order bias
  - Sampling bias
  - Biased reporting
- Sampling in qualitative research
  - Quota sampling
  - Purposive sampling
  - Theoretical sampling
  - Snowball sampling
  - Convenience sampling
- Generalizability in qualitative research
  - Sample-to-population generalization
  - Theoretical generalization
  - Case-to-case generalization = transferability

## Credibility in qualitative research

Credibility in qualitative research is an equivalent of internal validity in the experimental method. As you have seen, internal validity is a measure of the extent to which the experiment tests what it is intended to test. To ensure internal validity in experimental research we need to make sure that it is the IV, not anything else, that causes the change in the DV. To do this, we identify all the possible confounding variables and control them, either by

eliminating them or by keeping them constant in all groups of participants.

In a similar fashion, credibility in qualitative research is related to the question, “To what extent do the findings reflect the reality?” If a true picture of the phenomenon under study is being presented, the study is credible.

The term “**trustworthiness**” is also used to denote credibility in qualitative research.