SimplyPsychology

A-LEVEL PSYCHOLOGY REVISION NOTES

Research Methods

AQA Psychology 7181 (AS) and 7182 (A-level)

2025 specification · spec sections 3.2.3 (AS) and 4.2.3 (A-level)

Compiled by **Simply Psychology** · simplypsychology.org/research-methods.html

Contents

How to use these notes. Sections marked A-LEVEL are not assessed at AS. All other sections apply to both AS and A-level. Key terms are in **bold**; tables summarise strengths and limitations; "exam tips" call out the most common errors.

AQA 2025 SPECIFICATION — RESEARCH METHODS CONTENT

Methods

- **Experimental method** — laboratory and field experiments; natural and quasi-experiments.
- **Observational techniques** — naturalistic and controlled; covert and overt; participant and non-participant.
- **Self-report techniques** — questionnaires; structured and unstructured interviews.
- **Correlations** — analysis of the relationship between co-variables; difference between correlations and experiments.
- A-LEVEL Case studies and content analysis (including thematic analysis and coding).

Scientific processes

- **Aims and hypotheses** — stating aims; difference between aims and hypotheses; directional and non-directional hypotheses.
- **Sampling** — population vs sample; random, systematic, stratified, opportunity and volunteer; bias and generalisation.
- **Pilot studies** — and the aims of piloting.
- **Experimental designs** — repeated measures, independent groups, matched pairs.
- **Observational design** — behavioural categories; event sampling; time sampling.
- **Questionnaire construction** — open and closed questions; design of interviews.
- **Variables** — manipulation and control of IVs, DVs and extraneous variables; operationalisation.
- **Control** — random allocation, counterbalancing, randomisation, standardisation and control groups.
- **Demand characteristics and investigator effects.**
- **Ethics** — including the role of the British Psychological Society's code of ethics.
- **Peer review** — the role of peer review in the scientific process.
- **Implications of psychological research for the economy.**

Data handling and analysis

- **Quantitative and qualitative data** — and the distinction between qualitative and quantitative collection techniques.
- **Primary and secondary data** — including meta-analysis.
- **Descriptive statistics** — central tendency (mean, median, mode); dispersion (range, standard deviation); calculation of percentages; positive, negative and zero correlations.
- **Presentation of quantitative data** — graphs, tables, scattergrams, bar charts.
- **Distributions** — normal and skewed distributions and their characteristics.
- **Statistical testing — the sign test** — when to use and calculation.
- A-LEVEL Levels of measurement; inferential tests for difference and association; probability and significance; Type I and Type II errors.
- A-LEVEL Reporting investigations and features of science — objectivity, replicability, falsifiability, paradigms.

1 Experimental Method

An **experiment** is an investigation that looks for a **causal (cause-and-effect) relationship** between two variables. The researcher manipulates one variable (the **independent variable, IV**) and measures the effect on another (the **dependent variable, DV**) while controlling for other influences.

Figure 1.1 — In an experiment, the researcher manipulates the IV to create two or more conditions, which then influence the measured DV.

KEY TERMS

Independent variable (IV) — the variable that the researcher changes or manipulates to create the experimental conditions.

Dependent variable (DV) — the variable that the researcher measures; it is expected to change as a result of the IV.

WORKED EXAMPLES — OPERATIONALISED IVS AND DVS

Research question	Independent variable (IV)	Dependent variable (DV)
Does caffeine affect memory?	Caffeine dose: 0 mg vs 200 mg	Number of words correctly recalled from a 20-word list
Does sleep affect concentration?	Hours of sleep the night before: less than 5 vs at least 8	Time (seconds) to complete a Stroop task
Does background music affect revision?	Music condition: silence vs lyrics vs instrumental	Score (out of 20) on a multiple-choice quiz
Does group size affect conformity?	Number of confederates: 1, 2, 3 or 4	% of trials on which participant conforms to the wrong answer
Does mobile-phone use affect driving?	Mobile-phone condition: hands-free vs hand-held vs no phone	Reaction time (ms) to a brake-light stimulus in a driving simulator

Notice that each IV is **operationalised** (expressed in specific, measurable conditions) and each DV is a precise quantity that can be recorded objectively. Vague variables like "memory" or "concentration" become "number of words recalled" or "time on Stroop task".

Types of Experiment

Type	Description	Strengths	Limitations
Laboratory	Conducted in a controlled environment (not necessarily a literal lab) where extraneous variables can be carefully controlled. The researcher fully manipulates the IV.	High control of extraneous variables → high **internal validity**. Easily replicable due to **standardised procedures**.	Artificial setting → low **ecological validity**. Tasks may lack **mundane realism**. Risk of **demand characteristics**.
Field	Conducted in the participants' natural environment. The researcher still manipulates the IV.	Higher ecological validity than lab. Participants often unaware of being studied → reduces demand characteristics.	Less control of extraneous variables. Replication harder. Ethical issues (consent, debriefing) more difficult.
Natural	The IV is *not* manipulated by the researcher — it occurs naturally (e.g. effects of a natural disaster, deafness, a school's policy change).	Allows study of variables that could not be manipulated for ethical or practical reasons. High ecological validity.	Cannot randomly allocate participants → confounding variables. Rare events make replication difficult.
Quasi	The IV is a naturally-occurring difference between people (e.g. age, gender, having a mental disorder). The researcher does not — and cannot — manipulate it.	Allows comparisons that would be unethical to manipulate (e.g. clinical groups). Controlled DV measurement.	Lack of random allocation → potential confounding by participant variables. Cause-and-effect harder to establish.

EXAM TIP

The distinction between **natural** and **quasi** experiments is often confused. In a *natural* experiment the IV is an *event* (an earthquake, lockdown, school closure); in a *quasi* experiment the IV is a *characteristic of the participants* (age, gender, diagnosis). In both, the researcher cannot manipulate the IV.

2 Aims, Hypotheses and Variables

Aims

An **aim** is a general statement of what a researcher intends to investigate. Aims are derived from theory and previous research, and describe the *purpose* of the study.

> **EXAMPLE AIM**
> *"To investigate whether the amount of sleep the night before a memory task affects recall in sixth-form students."*

Hypotheses

A **hypothesis** is a precise, testable prediction about the relationship between the IV and the DV. A hypothesis must be **operationalised** — both variables must be expressed in measurable terms.

Type	When to use	Example
Directional (one-tailed)	Where previous research suggests the direction of effect.	"Participants who sleep *at least 8 hours* will recall **more** words on a 20-word list than participants who sleep *less than 5 hours*."
Non-directional (two-tailed)	Where previous research is mixed, ambiguous or absent.	"There will be a **difference** in the number of words recalled on a 20-word list between participants who sleep at least 8 hours and those who sleep less than 5 hours."
Null	Always written alongside the alternative hypothesis; states there is no effect.	"There will be **no difference** in the number of words recalled between the two sleep conditions; any apparent difference will be due to chance."

Variables

VARIABLES — AQA DEFINITIONS

Independent variable (IV) — the variable that is manipulated to see how it affects the dependent variable. For example, if a study investigates the effect of sleep on memory, the IV is the amount of sleep (e.g. 4 hours vs 8 hours).

Dependent variable (DV) — the variable that is measured. In the same example, the DV would be a measurement of memory, such as number of words recalled from a 20-word list.

Extraneous variable (EV) — any variable other than the independent variable that might affect the dependent variable. Researchers try to control or eliminate them.

Confounding variable (CV) — if an extraneous variable is not controlled or eliminated, and it does affect the DV, it becomes a confounding variable. This could lead to invalid conclusions about the effect of the IV on the DV.

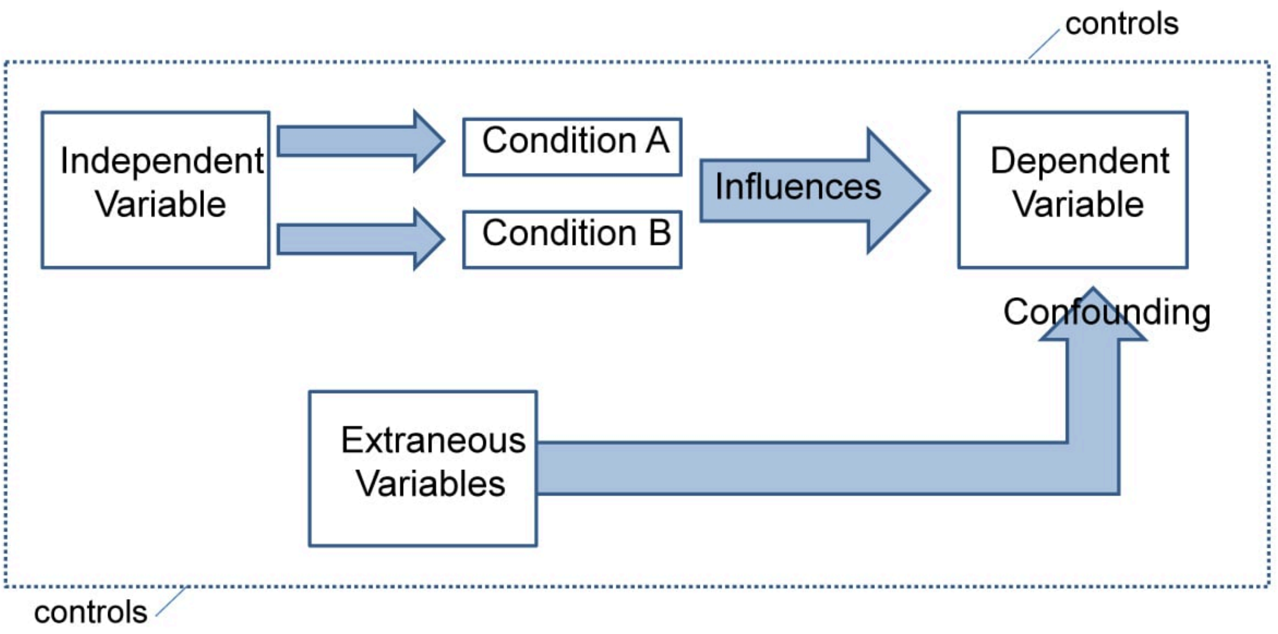


Figure 2.1 — A well-controlled experiment isolates the effect of the IV on the DV. ***Extraneous variables*** *that are not eliminated become* ***confounding variables****, biasing the apparent IV→DV relationship. The dashed border represents the controls a researcher uses to prevent this.*

Operationalisation

Operationalising variables is an important part of research design and hypothesis formulation. It involves **precisely identifying how any variable being studied will be measured or manipulated by the researcher**. Broad variables (sleep, memory, aggression) need to be broken down into quantifiable measures so the IV can be controlled and the DV can be measured.

NON-OPERATIONALISED VS OPERATIONALISED

Vague: "Aggression" → "How aggressive a child is."

Operationalised: "The number of times a child hits the Bobo doll in a 10-minute observation."

3 Sampling

A **sample** is the group of participants selected to take part in a study; they are drawn from the **target population**. A good sample is **representative** so that findings can be **generalised** back to the population.

Sampling Methods

Method	Procedure	Strengths	Limitations
Random	Every member of the target population has an equal chance of being selected — use a random-number generator or names-from-a-hat.	No researcher bias; theoretically the most representative method.	Need a full list of the population (sampling frame); volunteer dropout still possible.
Systematic	Select every *n*th person from a list (e.g. every 5th).	Objective; usually fairly representative.	Not truly random unless the starting point is randomised; pattern in the list could bias the sample.
Stratified	Identify subgroups (strata) in the population (e.g. age groups, gender). Select participants from each stratum in proportion to its size in the population.	Produces a representative sample of the strata identified; allows comparison between subgroups.	Identifying strata is time-consuming; representation of other unidentified differences may still be lacking.
Opportunity	Use whoever is available at the time the study is run (e.g. people in a corridor, sixth-formers in a free period).	Quick, easy and cheap.	Not representative; researcher bias in who is approached.
Volunteer (self-selected)	Participants opt in (e.g. responding to a poster or online advert).	Easy way to recruit motivated participants who tend to follow procedures carefully.	**Volunteer bias**: people who volunteer for research tend to differ from those who don't (more curious, more available).

EXAM TIP

When evaluating a sample, look first at **who is missing**. A study advertised on a university noticeboard recruits volunteers who are likely to be young, educated and curious — very different from the wider population. Generalisability is the question to ask, not just "is the sample big enough?"

4 Experimental Design

Experimental design describes *how participants are allocated to conditions* of the IV.

Design	Procedure	Strengths	Limitations
Independent groups	Two different sets of participants are used in the conditions of the study. Participants are allocated (usually randomly) to one condition only.	No **order effects** (no practice / boredom). Participants only see one condition → fewer demand characteristics.	**Participant variables** uncontrolled (the groups may differ in ability, motivation, age). Need more participants overall.
Repeated measures	The sample of participants all engage in all conditions of the study.	Participant variables are controlled (each participant is their own control). Fewer participants needed.	**Order effects** (fatigue, practice). Participants may guess the aim once they have done both conditions → demand characteristics.
Matched pairs	Different participants in each condition, but pairs are matched on key variables (age, gender, ability) before one of each pair is allocated to each condition.	No order effects; controls participant variables to some degree.	Time-consuming to match; never a perfect match; need a large pool to find matches.

Controlling for Order Effects: Counterbalancing

Counterbalancing is a technique used in **repeated-measures** designs to control for **order effects** (practice and fatigue). It does not *eliminate* order effects — instead, it spreads them **equally** across both conditions so they cancel out and cannot systematically bias the comparison between conditions.

> **WHY WE NEED COUNTERBALANCING**
>
> In a repeated-measures design, every participant completes *both* conditions. If everyone does condition A first and then B, any difference between the conditions could be caused by the IV — *or* by a **practice effect** (better at B because they've already done A) or a **fatigue effect** (worse at B because they're tired). These *order effects* become a confounding variable. Counterbalancing solves this by making sure the orders are balanced.

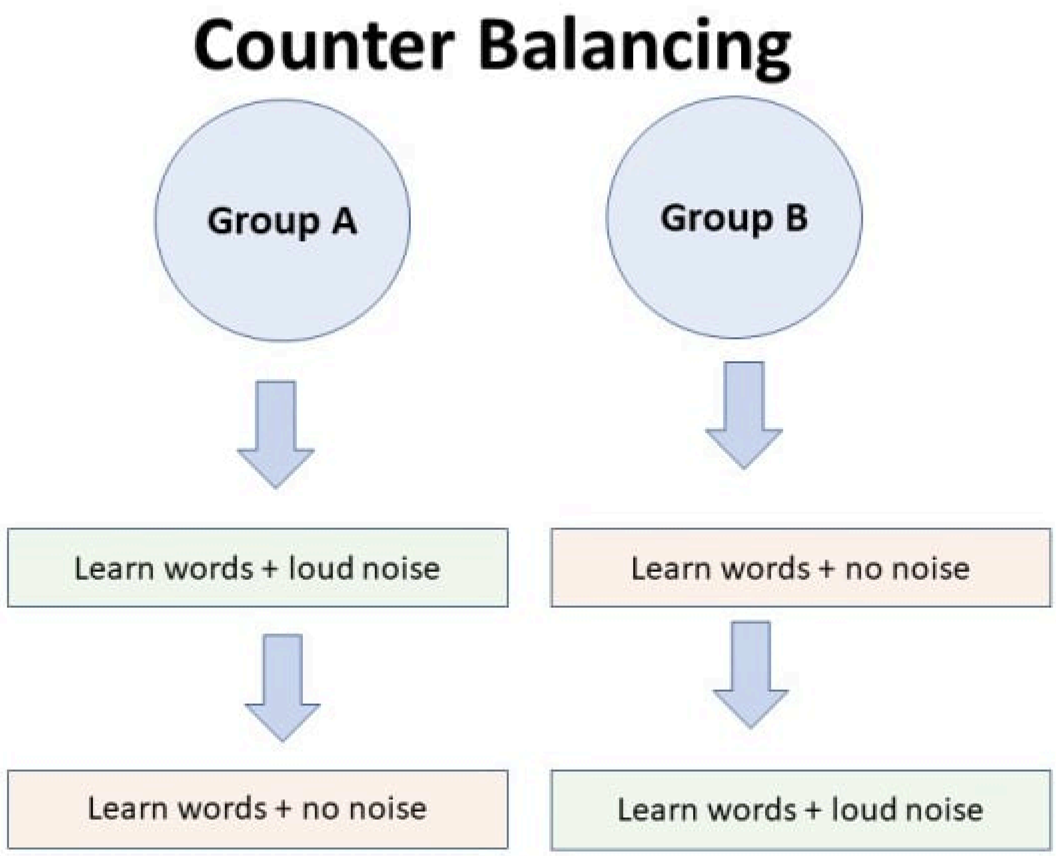


Figure 4.1 — Counterbalancing (AB / BA). The sample is split in half: Group 1 does condition A then B; Group 2 does B then A. Any practice or fatigue effect is therefore present in both conditions equally and cancels out in the overall comparison.

The two main counterbalancing techniques

Technique	How it works	When to use it
AB / BA (between-participants counterbalancing)	Split the sample in half. Half complete the conditions in the order **A** → **B**; the other half complete them in the order **B** → **A**. Order is balanced between groups.	The standard technique for two conditions. Easy to administer; produces a clean balance of practice/fatigue across conditions.
ABBA (within-participants counterbalancing)	*Every* participant completes all conditions twice, in the order **A** → **B** → **B** → **A**. The score for each condition is averaged across the two presentations.	Useful for very short tasks where doubling the procedure is feasible. Controls for order effects within each individual, not just across groups.

WORKED EXAMPLE — CAFFEINE AND REACTION TIME

A researcher tests whether caffeine improves reaction time. Each participant completes a reaction-time task **twice**: once after a placebo (condition A) and once after 200 mg of caffeine (condition B).

Without counterbalancing: All 30 participants do A then B. They are likely to score better on B simply because they have practised — the caffeine effect is confounded with a practice effect.

With AB / BA counterbalancing: 15 participants do A then B; the other 15 do B then A. Any practice effect now boosts condition B for the first group *and* condition A for the second group, so the practice effect appears equally in both conditions and cancels out. Any remaining difference can be attributed more confidently to the IV (caffeine).

Limitations of counterbalancing

- **It controls but does not eliminate order effects** — it just spreads them evenly. A very large practice effect can still increase overall noise in the data.
- **Not always possible** — for some studies the order is fixed by the procedure (e.g. a "before vs after" treatment study).
- **Carry-over effects can be asymmetric** — sometimes A affects B more than B affects A (e.g. learning a skill in A transfers to B but not vice versa). In these cases counterbalancing cannot fully cancel the effect.
- **More than two conditions becomes complex** — with three conditions there are six possible orders (ABC, ACB, BAC, BCA, CAB, CBA); fully balanced designs require larger samples. Researchers often use a **Latin square** to manage this.

EXAM TIP — COUNTERBALANCING VS RANDOM ALLOCATION

Students often confuse the two. **Counterbalancing** is used in *repeated-measures* designs to control *order effects*. **Random allocation** is used in *independent-groups* designs to control *participant variables*. They solve different problems and apply to different designs.

Random Allocation

In independent-groups designs, participants are **randomly allocated** to conditions (e.g. by coin flip or random number) so that participant variables are balanced across groups by chance.

ORDER EFFECTS

Practice effect — performance improves on the second condition because participants have learned the task.

Fatigue effect — performance worsens on the second condition because participants are tired or bored.

5 Pilot Studies and Control Techniques

Pilot Studies

A **pilot study** is a small-scale trial run of a study, conducted before the main investigation, to identify problems with the design before committing resources to the full study.

Aims of piloting

- Check that the procedure is clear and standardised.
- Check that the materials work (e.g. that a word list is the right length and difficulty).
- Identify **floor effects** (task too hard — everyone scores zero) or **ceiling effects** (task too easy — everyone gets full marks).
- Trial the timings, instructions and debriefing.
- For self-report studies: test whether questions are clear, neutral and produce useful answers.

Standardisation

Standardisation means using exactly the same procedures and instructions for every participant. This eliminates differences in how participants are treated as a source of extraneous variation, raising the **reliability** of the study.

Randomisation

Randomisation uses chance to reduce the influence of bias on the design (e.g. randomising the order of stimuli, random allocation to conditions, randomising who is asked first).

Control Groups and Conditions

A **control group** does not receive the experimental treatment. It provides a baseline against which the experimental group can be compared — without it, the researcher cannot tell whether the change in the DV was caused by the IV or by something else (the placebo effect, time passing, etc.).

> **WHY A CONTROL MATTERS**
>
> A drug trial measures mood before and after a new antidepressant. Patients report improved mood — but is it the drug? Without a control group taking a placebo, we cannot tell whether the improvement is due to the active ingredient or to the patients *expecting* to feel better.

Single-Blind and Double-Blind Procedures

Single-blind: the participant doesn't know which condition they are in. Reduces participant bias and demand characteristics.

Double-blind: *neither* the participant nor the researcher administering the task knows which condition the participant is in. Eliminates both participant bias AND investigator effects. The gold standard in drug trials.

6 Demand Characteristics and Investigator Effects

Demand Characteristics

Demand characteristics are any features of a research situation that give the participant clues about the aim of the study, leading them to alter their behaviour from how they would normally act.

Two main effects:

- **"Please-You" effect** — participants try to *help* the researcher by giving the responses they think are wanted.
- **"Screw-You" effect** — participants deliberately give the opposite of what they think is wanted to spoil the study.

How to reduce demand characteristics:

- Use a **single-blind procedure**.
- Use an **independent-groups design** so participants only see one condition.
- Use a **cover story** that disguises the true aim.
- Use a **field experiment** where participants don't know they are being studied.

Investigator Effects

Investigator effects are any unwanted influence of the researcher's behaviour — conscious or unconscious — on the DV. Includes:

- Subtle differences in how the experimenter behaves in different conditions (tone of voice, body language).
- The way participants are selected for conditions.
- Biased interpretation of ambiguous data.
- Researcher's expectations leaking into the procedure (Rosenthal effect).

How to reduce investigator effects:

- Use a **double-blind procedure**.
- **Standardise** instructions and procedures.
- Use **randomisation** in the design.
- Have **multiple observers** and use **inter-rater reliability** checks (see Section 12).

EXAM TIP

Demand characteristics threaten **internal validity** (the cause-and-effect link). Investigator effects threaten both **internal validity** (the result may be due to the researcher, not the IV) AND **reliability** (different researchers may produce different results). Be precise in your wording.

7 Observational Techniques

Observation involves watching and recording behaviour. It is *not* an experiment because the researcher does not manipulate an IV — they only measure existing behaviour. Observations can complement experiments by providing a richer picture.

Types of Observation

Dimension	Type	Description
Setting	Naturalistic	Behaviour observed in the participants' natural environment. High ecological validity; less control of extraneous variables.
	Controlled	Behaviour observed in a structured (often lab) setting. More control; lower ecological validity (artificial).
Researcher visibility	Overt	Participants *know* they are being observed. Ethically straightforward but introduces demand characteristics.
	Covert	Participants *do not know* they are being observed. Reduces demand characteristics; raises ethical concerns about consent and privacy.
Researcher involvement	Participant	The researcher is part of the group being observed (e.g. living with a community for a year).
	Non-participant	The researcher remains separate from the group ("fly on the wall").

Observational Design

Behavioural categories

Behavioural categories break the behaviour to be observed into discrete, observable, mutually exclusive units. For example, when observing a child playing:

- Solitary play (playing alone)
- Parallel play (playing alongside but not with)
- Cooperative play (taking turns, sharing, joint activity)
- Onlooker behaviour (watching others play but not joining in)

Each category must be operationalised so that two observers would code the same behaviour the same way.

Event sampling and time sampling

Event sampling: every instance of a target behaviour is recorded throughout the observation period. Useful for infrequent behaviours.

Time sampling: the observer records what behaviour is occurring at fixed intervals (e.g. every 30 seconds). Useful when behaviours are frequent or multiple behaviours occur at once.

EXAM TIP

When asked to suggest behavioural categories for a scenario, give *three* categories that are **operationalised** (specific, observable) and **mutually exclusive** (no overlap). Avoid vague categories like "happy" or "behaving well".

8 Self-Report Techniques

Self-report methods ask participants to give information about themselves — their thoughts, feelings, behaviours, attitudes. The two main techniques are **questionnaires** and **interviews**.

Questionnaires

A **questionnaire** is a written set of questions delivered on paper or online.

Open and Closed Questions

Type	Format	Data produced	Strength	Limitation
Closed	Fixed responses (Yes/No, multiple choice, Likert scale, ranking).	Quantitative	Easy to analyse statistically; participants find them quick to answer.	Limits the range of responses; depth/nuance lost.
Open	No fixed responses — participants answer in their own words.	Qualitative	Captures rich detail and unexpected responses.	Harder to analyse and compare; subjective interpretation needed.

Interviews

Type	Description	Strengths	Limitations
Structured	A predetermined list of questions delivered in the same order to every participant.	High **reliability** — replicable. Easy to compare responses across participants.	No follow-up of interesting answers; rigid format may miss important detail.
Unstructured	Loose, conversational; the interviewer follows interesting threads as they emerge.	Rich, in-depth **qualitative** data. Can probe unexpected responses.	Hard to replicate; risk of interviewer bias; difficult to compare responses.
Semi-structured	A list of core questions plus freedom to probe answers further. Common in clinical and qualitative research.	Balances depth (unstructured) with comparability (structured).	Requires skilled interviewer; analysis is time-consuming.

Designing Good Self-Report Items

- **Avoid leading questions** ("Don't you agree that…?") — they bias responses.
- **Avoid double-barrelled questions** ("Do you like cats and dogs?") — participants can't answer cleanly if their view differs for cats vs dogs.

- **Avoid jargon** — use accessible language.
- **Pilot** the questionnaire to spot ambiguities and refine wording.
- **Guarantee anonymity** to reduce **social desirability bias** (the tendency to give the "acceptable" answer rather than the truthful one).

9 Correlations

A **correlation** measures the strength and direction of a relationship between two **co-variables** (variables that are measured, not manipulated).

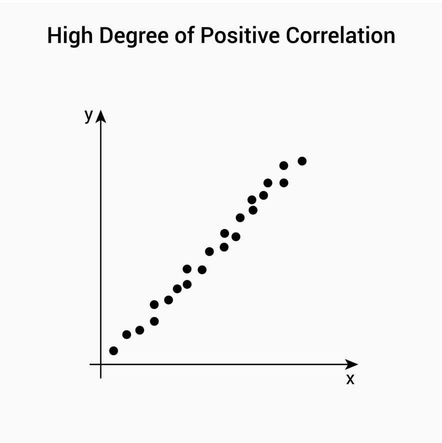
KEY TERMS

Co-variables — the two variables that are measured in a correlational study. There is no manipulation of variables in correlational studies, so there are no independent or dependent variables, only co-variables. This means no causal conclusions can be drawn — correlation does not infer causation.

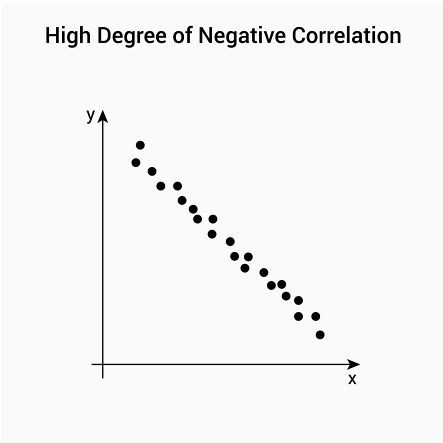
Correlation coefficient — a value calculated from the data that shows the strength and direction of the relationship between co-variables. Values range from –1 to +1. **Pearson's r** is used with interval-level data; **Spearman's rho** can be used with interval or ordinal data. A value of −1 = perfect negative linear relationship; +1 = perfect positive; 0 = no relationship. The closer to ±1, the stronger.

Types of Correlation

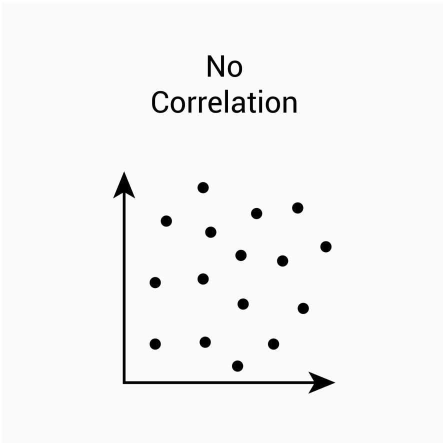
Type	Coefficient	Interpretation
Positive	+0.1 to +1.0	As one variable increases, the other tends to increase. E.g. hours of sleep × test scores.
Negative	–0.1 to –1.0	As one variable increases, the other tends to decrease. E.g. stress level × immune function.
Zero	≈ 0	No systematic relationship between the variables.



Positive correlation
r ≈ +0.85



Negative correlation
r ≈ −0.85



Zero / no correlation
r ≈ 0

Strength of correlation (interpreting r)

|r| value	Strength
0.00 – 0.19	Very weak / none
0.20 – 0.39	Weak
0.40 – 0.59	Moderate
0.60 – 0.79	Strong
0.80 – 1.00	Very strong

Differences Between Correlations and Experiments

Feature	Experiment	Correlation
Variable status	IV manipulated; DV measured	Both variables measured
Causal claims	Can infer cause-and-effect	Cannot infer cause-and-effect
Random allocation	Required for cause inference	Not applicable
Data display	Bar charts, line graphs	Scattergrams

EXAM TIP — CORRELATION ≠ CAUSATION

A correlation only shows that two variables vary together — it does **not** show that one causes the other. A third variable may explain both (the "third variable problem"), or the causation may run in the opposite direction. Always include this caveat when interpreting correlations.

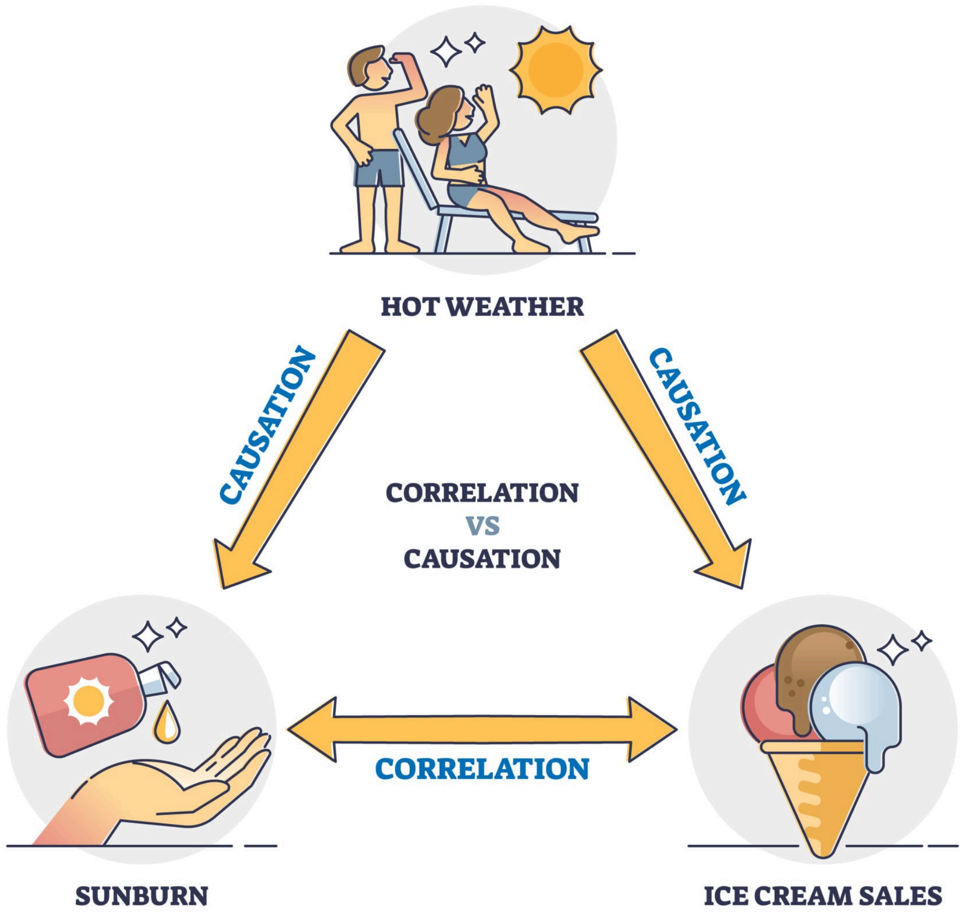


Figure 9.1 — A classic illustration of correlation vs causation. ***Sunburn*** *and* ***ice cream sales*** *are positively correlated, but neither causes the other. The true cause of both is* ***hot weather*** *— a third variable that influences both co-variables.*

10 Case Studies and Content Analysis

A-LEVEL

Case Studies

A **case study** is an in-depth investigation of a single individual, small group, organisation or event — typically over an extended period and using multiple methods (interviews, observations, psychometric tests, archival records).

Famous case studies in psychology

- **HM (Henry Molaison)** — surgical removal of the hippocampus to treat epilepsy left HM unable to form new long-term memories; he was studied for decades. Revealed the role of the hippocampus in long-term memory.
- **Phineas Gage** — railroad worker who survived a tamping iron passing through his frontal lobes (1848); dramatic personality changes informed understanding of frontal-lobe function.
- **Little Hans** — Freud's case study of a five-year-old's horse phobia, used to develop the Oedipus complex.
- **Genie** — a child kept in extreme social isolation until age 13; provided rare evidence about critical periods in language acquisition.

Aspect	Strengths	Limitations
Data richness	Rich, detailed qualitative data; can capture rare or unique phenomena that would be impossible to study experimentally.	Findings cannot be generalised to a wider population — a sample of one.
Ecological validity	Studies the person in their real-life context.	Researcher subjectivity in selecting and interpreting evidence is hard to avoid.
Hypothesis-generating	Famous cases (HM, Gage) have transformed entire research fields.	Cannot establish cause-and-effect; conclusions are correlational at best.

Content Analysis

Content analysis is a method for converting qualitative material (TV shows, newspaper articles, diaries, interview transcripts) into **quantitative data** by counting the occurrences of pre-defined categories.

The procedure

1. **Define a research question** (e.g. "How are women portrayed in evening commercials?").
2. **Decide on coding categories** in advance (e.g. *parent, professional, romantic partner, sexual object, comedy figure*) that are mutually exclusive and clearly defined.
3. **Sample** the material — e.g. 40 hours of evening television across three channels.

4. **Code** each instance into its category.
5. **Quantify** — count frequencies, calculate percentages, display as bar charts.
6. **Check reliability** by having a second coder independently code an overlapping subset and correlating the two sets of counts (**inter-rater reliability**).

> **CODING IN CONTENT ANALYSIS**
>
> **Coding** means assigning each unit of qualitative material to a category. Good codes are **mutually exclusive**, **exhaustive**, and **operationalised** so any trained coder would assign the same unit to the same category.

11 Ethics in Research

The **British Psychological Society (BPS) Code of Ethics** sets out the principles psychologists must follow when conducting research with human participants. Researchers must weigh the **benefits** of the research against the potential **costs** to participants.

Core Ethical Issues

Issue	Description	How researchers deal with it
Informed consent	Participants must be told what the study involves before agreeing to take part.	Provide a written information sheet and signed consent form. For under-16s, parent/guardian consent is also required.
Deception	Participants should not be deliberately misled about the aim or procedure.	Use the minimum necessary deception; debrief fully afterwards; give participants the chance to withdraw their data.
Right to withdraw	Participants must be free to leave the study at any point, without penalty.	State this clearly at the start; remind participants during the study; honour requests to withdraw data after the study.
Protection from harm	Physical and psychological harm should not exceed the risks of everyday life.	Pilot the study to identify risks; stop if a participant becomes distressed; provide follow-up support.
Confidentiality	Personal data must be protected (Data Protection Act / GDPR).	Use participant numbers, not names; store data securely; report aggregate results only.
Privacy	Participants have the right not to be observed where they would expect privacy.	Avoid covert observation in private spaces; gain consent for sensitive data collection.
Debriefing	Participants must be returned to their pre-study state and given the chance to ask questions.	Provide a written or spoken debrief; explain any deception; remind participants of their right to withdraw data.

EXAM TIP — ETHICS IN SCENARIOS

When an exam question asks you to identify an ethical issue in a study, name the **specific issue** (e.g. "lack of informed consent") rather than the generic "it was unethical". Then explain **how the researcher could deal with it** in that scenario.

12 Reliability and Validity

Reliability

Reliability refers to the **consistency** of a measurement. A reliable measure produces similar results when repeated under the same conditions.

Type	What it checks	How to measure
Test–retest	Whether the same test produces similar scores on different occasions.	Administer the test twice (with a gap); correlate the two sets of scores. A correlation of **r > 0.80** indicates good reliability.
Inter-rater (inter-observer)	Whether different researchers code the same data consistently.	Two or more coders independently rate the same material; correlate their judgements. r > 0.80 indicates good reliability.
Internal	Whether all the items in a test measure the same underlying construct.	Split-half reliability: correlate scores on the first half with the second half of a test.

Improving reliability

- Use **standardised procedures** for all participants.
- **Train observers** and pilot coding schemes before the main study.
- Use **operationalised** behavioural categories with clear definitions.
- For questionnaires, refine ambiguous items based on pilot results.

Validity

Validity refers to **whether a study measures what it claims to measure**, and whether the findings reflect what is true in the real world.

Type	Question it asks
Face validity	Does the test *look like* it measures the construct? (Subjective.)
Concurrent validity	Does it correlate with an established measure of the same construct?
Internal validity	Did the IV (not extraneous variables) cause the change in the DV?
External validity	Can the findings be generalised beyond the study?
Ecological validity	Can the findings be generalised to *other settings* (real-world contexts)?
Population validity	Can the findings be generalised to *other people* (beyond the sample)?
Temporal validity	Can the findings be generalised to *other time periods* (cultural or historical contexts)?

Improving validity

- Use a **double-blind procedure** to reduce investigator effects.
- Use a **control group** to rule out placebo effects.
- **Operationalise** variables precisely so the study measures what it claims.
- Use **diverse samples** to improve population validity.
- Conduct **field studies** alongside lab studies to check ecological validity.

13 Data Handling and Analysis

Quantitative and Qualitative Data

Type	Description	Strengths	Limitations
Quantitative	Numerical data — counts, scores, measurements.	Easy to analyse statistically; allows comparison across groups; objective.	May miss nuance and context; reduces complex experiences to numbers.
Qualitative	Descriptive data — words, themes, narrative.	Captures rich detail and meaning; useful for exploratory research.	Subjective interpretation; harder to compare and generalise.

Primary and Secondary Data A-LEVEL

Primary data is information collected first-hand by the researcher for the current study. **Secondary data** is information already collected for another purpose that is then re-analysed (e.g. government statistics, published research findings).

Meta-analysis

A **meta-analysis** statistically combines the findings of many studies on the same research question to produce a single, pooled effect size. It uses **secondary data** and gives much greater statistical power than any single study.

	Strengths	Limitations
Meta-analysis	Large combined sample → high statistical power and generalisability; less influence of any single outlier study.	**Publication bias / file-drawer problem** — only significant findings tend to be published; **heterogeneity** — studies may differ in method and sample.

Descriptive Statistics

Measures of central tendency

Measure	How to calculate	Best for
Mean	Sum of all values ÷ number of values.	Interval/ratio data with no outliers.
Median	Middle value when data is ranked.	Ordinal data or data with outliers.
Mode	Most frequently occurring value.	Nominal (categorical) data.

Measures of dispersion

- **Range** = highest value – lowest value. Simple but sensitive to outliers.
- **Standard deviation (SD)** = the average distance of values from the mean. More informative; uses all data points; expressed in the same units as the data.

Presentation of Data

Graph	Use
Bar chart	Frequencies of **categorical** (discrete) data. Bars are **separated**.
Histogram	Frequencies of **continuous** data grouped into intervals. Bars **touch**.
Scattergram	Relationship between two co-variables (correlation).
Line graph	Continuous data where the x-axis order matters (e.g. time series).
Table	Exact numerical values for reference.

Distributions

The **normal distribution** is a symmetrical bell-shaped curve. The mean, median and mode all sit at the centre. Most scores cluster near the mean and become rarer further away.

A distribution is **positively skewed** when most scores are low and a few high scores stretch the right tail (e.g. income). It is **negatively skewed** when most scores are high and a few low scores stretch the left tail (e.g. exam scores where the test was easy).

Levels of Measurement A-LEVEL

Level	What it means	Example
Nominal	Categorical — labels with no order.	Eye colour (blue/brown/green), gender.
Ordinal	Ordered, but unequal intervals.	Race finishing positions; Likert scale agreement.
Interval	Equal intervals; arbitrary zero.	Temperature in °C (0°C ≠ no temperature).
Ratio	Equal intervals; true zero.	Reaction time in ms; weight in kg.

14 Inferential Testing A-LEVEL

Inferential statistics allow us to draw conclusions about a population based on a sample. We test whether the difference (or correlation) observed in our data is large enough to be unlikely due to chance.

Probability and Significance

The standard significance level in psychology is **$p \leq 0.05$** — the probability that the observed result is due to chance is 5% or less.

If **$p \leq 0.05$** → reject the null hypothesis; the result is **significant**.

If **$p > 0.05$** → retain the null hypothesis; the result is **not significant**.

Type I and Type II errors

- **Type I error (false positive)** — rejecting the null hypothesis when it is actually true. More likely when significance level is too lenient (e.g. $p \leq 0.10$).
- **Type II error (false negative)** — retaining the null hypothesis when it is actually false. More likely with a too-strict significance level or a small sample.

The Sign Test

The **sign test** is a simple non-parametric test for differences in **related** (repeated-measures or matched-pairs) designs with **nominal-level** data.

When to use the sign test (three conditions)

1. The study tests for a **difference** (not a correlation).
2. The design is **related** — same participants in both conditions, or matched pairs.
3. The data is at the **nominal level** — each pair can be reduced to "+", "–", or "no change".

How to calculate the sign test

1. For each participant, calculate the difference between scores in the two conditions.
2. Record the **sign** of each difference (+, –, or 0).
3. **Discard the zeros**; this is now your effective *N*.
4. Count the less frequent sign — this is **S** (the test statistic).
5. Look up the critical value of S for your N and significance level in the sign-test critical-values table.
6. **If S ≤ critical value, reject the null hypothesis** — the result is significant.

EXAM TIP — SIGN TEST DIRECTION

The sign test reverses the usual direction: you reject H_0 when calculated S is **less than or equal to** the critical value (not greater than). Many students forget this. Remember: *"S small → significant"*.

Choosing a Statistical Test A-LEVEL

	Nominal data	Ordinal data	Interval/ratio data
Difference, related design	Sign test	Wilcoxon	Related t-test
Difference, independent design	Chi-squared (χ^2)	Mann–Whitney U	Unrelated t-test
Association / correlation	Chi-squared (χ^2)	Spearman's rho (ρ)	Pearson's r

MNEMONIC — CHOOSING A TEST

Three questions: (1) **What are you testing?** (difference or association) (2) **What design?** (related or independent) (3) **What level of data?** (nominal, ordinal, or interval/ratio). The intersection in the table gives you the test.

15 Peer Review, Reporting and Features of Science A-LEVEL

Peer Review

Peer review is the process by which research is assessed by experts in the same field before it is published in an academic journal. Reviewers evaluate the study's methodology, originality, validity, and clarity, and recommend whether it should be accepted, revised or rejected.

Functions of peer review

- **Quality control** — filters out methodologically flawed or fraudulent research.
- **Allocation of research funding** — funding bodies use peer review to decide which proposals to support.
- **Knowledge validation** — gives the wider scientific community confidence in published findings.

Limitations of peer review

- **Publication bias** — significant or novel results are more likely to be published, while null results are filed away. This distorts the published literature.
- **Bias toward established researchers and theories** — reviewers may favour work that agrees with their own.
- **Slow** — peer review can take months or years.
- **Anonymity** can encourage reviewers to be harsher or to dismiss work they disagree with.

Extended evaluation (AO3) — peer review

Strength — protects the integrity of science. A major strength of peer review is its role as a quality-control gatekeeper. Reviewers identify methodological flaws, fraudulent data and weak conclusions before findings enter the published literature — for example, Wakefield's 1998 MMR–autism paper was eventually retracted only after persistent post-publication scrutiny, illustrating why front-end review matters. This is important because, without peer review, flawed research could be cited as evidence in clinical and policy contexts. This strengthens public trust in psychology and supports its claim to be a rigorous, evidence-based discipline.

Limitation — publication bias and the file-drawer problem. However, a major limitation is publication bias. Reviewers and journal editors disproportionately favour statistically significant and novel findings, leaving null results unpublished (the "file-drawer problem"). This is important because meta-analyses and theoretical reviews are then built on a skewed evidence base — effect sizes are over-estimated, and replications that fail to reach significance never appear. This weakens the validity of the published literature and limits peer review's ability to deliver a fully representative picture of psychological research.

Limitation — bias against novel or critical work. A further limitation is that peer review can be conservative. Reviewers are typically senior figures in a field who may favour work consistent with established theories and disfavour findings that challenge their own positions. Smith (1999) found low inter-rater reliability between reviewers of the same paper, suggesting subjective judgement plays a large role. This is important because genuinely paradigm-shifting work may be filtered out, slowing scientific progress. This limits peer review's effectiveness as a neutral quality check and supports the case for open review and pre-registration.

Application — economy and policy implications. Peer review has clear applied value. Funding bodies (e.g. the ESRC, MRC, Wellcome Trust) use peer review to allocate scarce research funding, ensuring public money supports rigorous work; NICE and NHS guidance relies on peer-reviewed evidence to recommend treatments. This is important because credible filtering of evidence prevents wasteful funding of weak research and supports cost-effective public-health policy. This is a clear example of how the scientific process delivers measurable economic benefit, even if its limitations are real.

Implications of Psychological Research for the Economy

Psychological research has real-world economic implications:

- Research on **attachment** has informed shared parental leave policy — both parents can take leave because the father can also be a primary attachment figure.
- Research on **mental health treatments** (CBT, drugs) has informed NHS treatment guidelines, reducing the economic cost of untreated mental illness (lost productivity, healthcare burden).
- Research on **eyewitness testimony** has informed police interview protocols, reducing wrongful convictions.
- Research on **workplace stress** has informed Health and Safety Executive guidance, reducing sickness absence costs.

Reporting Psychological Investigations

A research report typically contains the following sections (in this order):

1. **Abstract** — a 150–250 word summary of aims, methods, results and conclusions.
2. **Introduction** — background literature, theoretical context, leading to the aim and hypothesis.
3. **Method** — design, participants, materials, procedure (in enough detail to allow replication).
4. **Results** — descriptive and inferential statistics, tables, graphs.
5. **Discussion** — interpretation of findings, links to previous research, limitations, implications.
6. **Referencing** — full APA-style references for all sources cited.

Features of Science

Objectivity and the empirical method

Science aims to be **objective** — free from researcher bias — and **empirical** — based on observation and measurement rather than belief.

Replicability and falsifiability

Replicability: a finding must be repeatable by independent researchers. **Falsifiability** (Popper): a scientific theory must make predictions that could in principle be disproved. Unfalsifiable theories (such as some psychoanalytic claims) are not scientific.

Theory construction and hypothesis testing

Science proceeds by formulating theories, deriving testable hypotheses, conducting empirical tests, and revising theories in light of the results.

Paradigms and paradigm shifts

Kuhn (1962): science operates within **paradigms** — shared frameworks of assumptions, methods and theories. Occasionally, accumulating anomalies lead to a **paradigm shift** — a major theoretical revolution (e.g. behaviourism → cognitive revolution; Newtonian → relativistic physics).

These revision notes were prepared for Simply Psychology and cover spec sections 3.2.3 (AS) and 4.2.3 (A-level) of the AQA Psychology 2025 specification. For deeper coverage of any topic, see the corresponding article at simplypsychology.org/research-methods.html.