

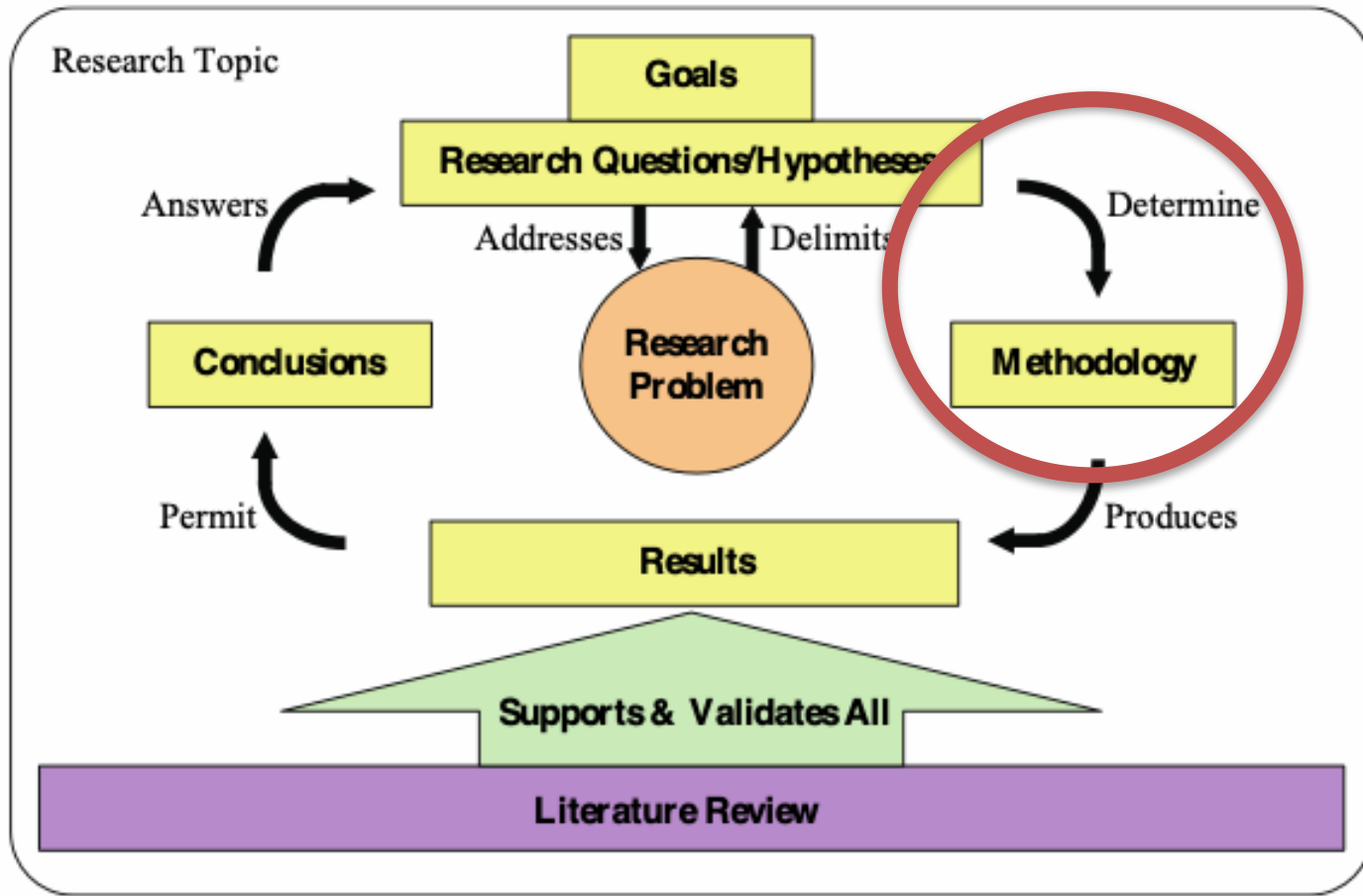


Preparing a research project

Research designs and practicalities

Mits Ota

Determining methodology



3 dimensions of methodology

Design type

Experimental

Correlational

Modelling

etc.

Grouping

Cross-sectional

Longitudinal

Treatment

etc.

Data type

Behavioural

Neural

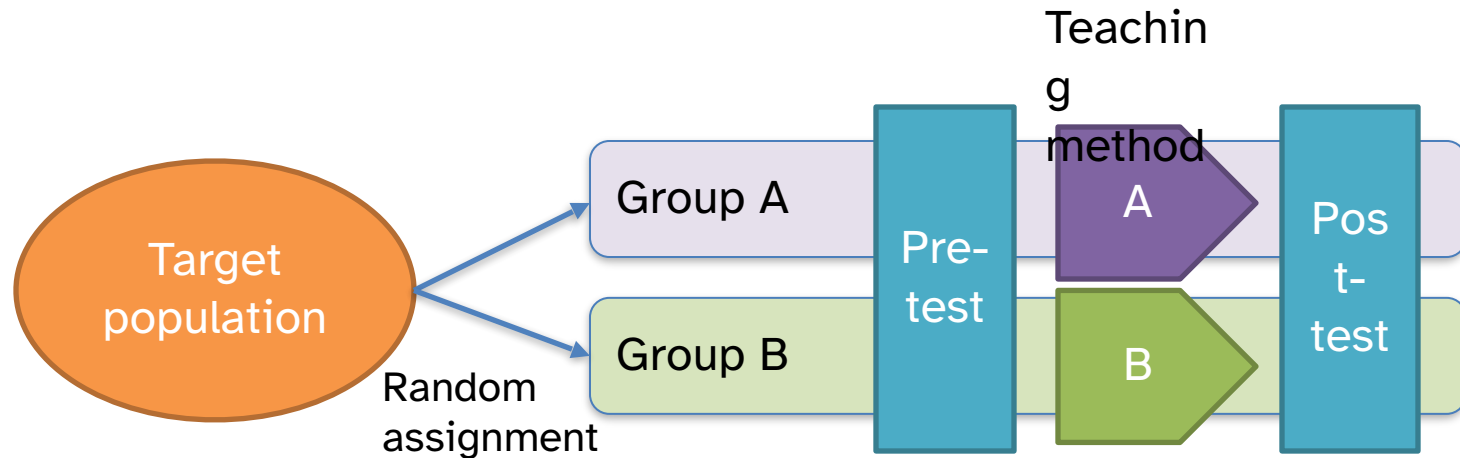
Corpus

etc.

Design types

Experimental studies

- Manipulate (i.e., systematically vary) the independent variable and control extraneous variables.
- Strong case for causality, but potential issues with validity.



Design types

Correlational (ex post facto) studies

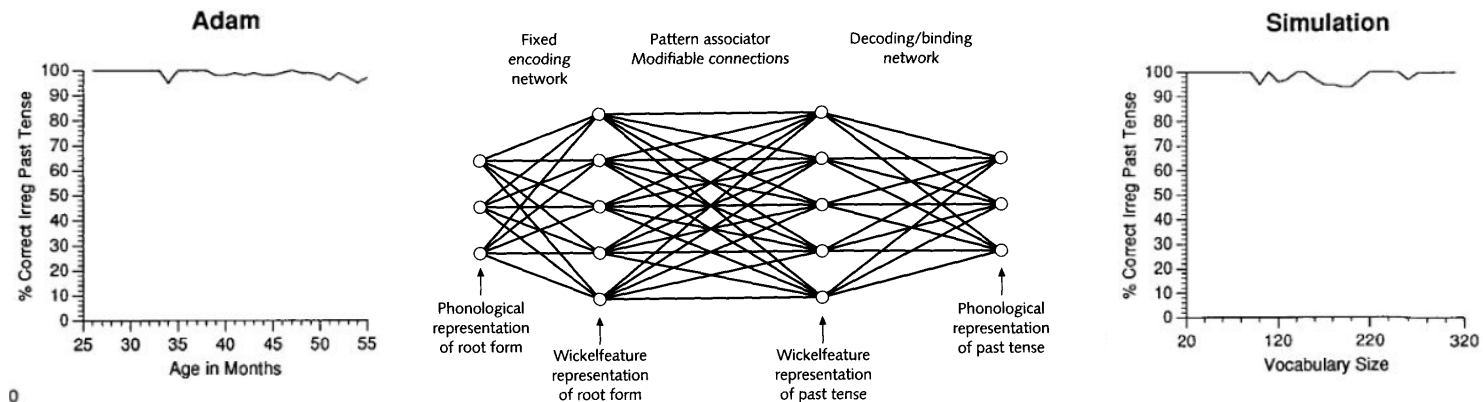
- Assess the statistical relationship between variables without manipulation or control.
- Causality cannot be established; but the design allows us to explore ‘what is going on’; often used with corpus data.

Learner	Method	Sex	Age	L1	Degree	WM	L2 level
Abby	A	F	17	En	UG	7.3	6.8
Björn	B	M	37	Sw	UG	4.9	7.2
Cindy	B	F	30	En	PG	5.4	8.4
Dave	A	M	25	En	HS	6.6	7.3
Edith	B	F	23	Fr	UG	7.0	9.1
...

Design types

Modelling (simulation) studies

- Examine whether patterns in real data can be simulated by a simplified model that represents a system (e.g., learning).
- Can substitute direct experimentation.



Grouping

Cross-sectional

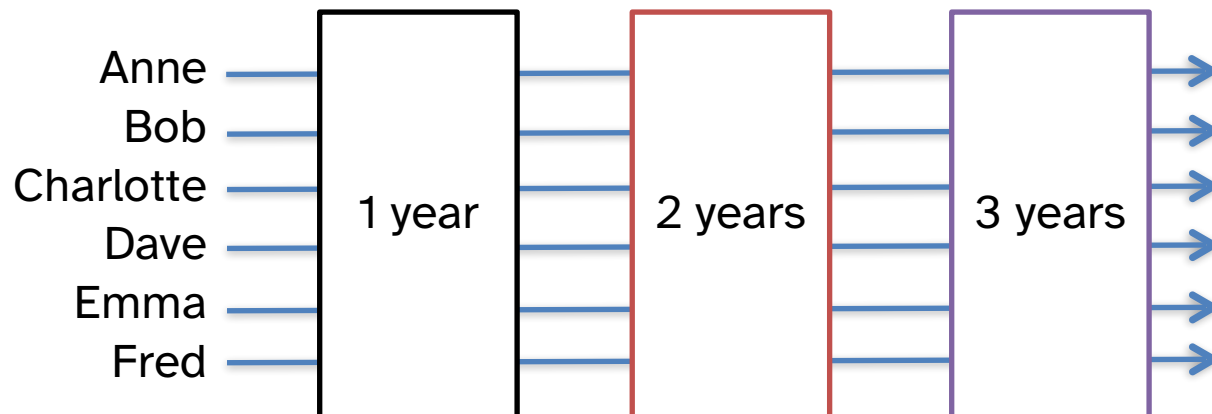
- Samples of predefined populations (e.g., age, language)
- No carry-over effects; but more noise from individual differences



Grouping

Longitudinal

- Data collection over time from same participants
- Carry-over effects; but less variation noise
- Typically fewer participants than cross-sectional design due to constraints on recruitment and attrition



Grouping

Experimental

- Randomly assign a pool of participants to different experimental groups/conditions
- Or collect different types of data from the same participants: e.g., before and after a ‘treatment’ (i.e., pre-test/post-test)

Intact

- Pre-assigned groups of non-inherent characteristics: e.g., participants in Class A versus Class B

Data type

Behavioural

- On-line data: Real-time measurements of participants' performance: e.g., reaction time, eye-movement
- Off-line data: Untimed measurements of participants' performance: e.g., grammatical judgments, comprehension
- Elicited production data

Neural

- Measurements of brain activities in response to stimuli: electrical signals (electroencephalogram; EEG) or blood flow (functional magnetic resonance imaging; fMRI, optical tomography; OT)

Data type

Corpus

- Recorded and/or transcribed linguistic data of spontaneous or structured speech
- Learner corpora: CHILDES, TalkBank, ICLE

Self-reported data

- Information about participants reported by themselves: e.g., questionnaires/surveys, self-rating of language proficiency, vocabulary list (MacArthur CDI)

Selecting designs: An example

What methods can be used to address this research question?

When do English-speaking children acquire the passive construction?

- Independent factor: Age of children
- Dependent factor: However you operationalise the 'acquisition' of passives in your design/data type

Selecting designs: An example

Method 1: Corpus analysis

- Longitudinal data of spontaneous production
- Data: First recorded cases of non-imitated passives
- IV = age; DV = first case



```
6496 *CHI:      they won't hurt <their [?]> head by icicles . ▶
6497 %mor:      pro:sub|they mod|will-neg|not v|hurt&ZERO det:poss|their n|head
6498          prep|by n|icicle-PL .
6499 %gra:      1|4|SUBJ 2|4|AUX 3|2|NEG 4|0|ROOT 5|6|MOD 6|4|OBJ 7|4|JCT 8|7|POBJ
6500          9|4|PUNCT
6501 %pho:      der 'wʊ 'hʌ də 'he bɑɪ 'aɪsɪɡlθ
6502 %mod:      'ðeɪ 'wɒnt 'hʌt 'ðeɪ 'hed 'bɑɪ 'aɪsɪkəlz
```

▶ 00:00 / 03:24 🔊 🔍 📄

Selecting designs: An example

Method 2: Priming experiment

- Cross-sectional groups: Age 2, 3 and 4
- Data: Children's elicited production
- IV = age; DV = active vs passive response



“Hi Thomas! Hi Percy!”



“Look! Percy is pushed
by Thomas”



“What’s happening
now?”

Selecting designs: An example

Method 3: Preferential looking

- Cross-sectional groups: Age 2, 3 and 4
- Data: Eye movement between two images
- IV = age; DV = looking time to matching picture



Look! Percy is pushed by Thomas.

Important design concepts

Between vs within-subjects (repeated-measures) design

	Participant	Condition
Group A	Andy	X
	Beth	X
	Charles	X
	Debbie	X

	Participant	Condition
Group B	Eddie	Y
	Frances	Y
	George	Y
	Hannah	Y

Participant	Condition	
Andy	X	Y
Beth	X	Y
Charles	X	Y
Debbie	X	Y
Eddie	X	Y
Frances	X	Y
George	X	Y
Hannah	X	Y

Example: X & Y = word-learning methods

Important design concepts

Between vs within-subjects (repeated-measures) design

Between-subjects

- ‘Noisy’ and less powerful
- No carry-over effects

Within-subjects

- Participants held constant across conditions
- More statistically ‘powerful’
- Carry-over effects; not all conditions can be repeated
- Counter-balancing necessary

Important design concepts

Factorial design

Two or more factors and their interactions can be tested by including all permutations of the levels of factors

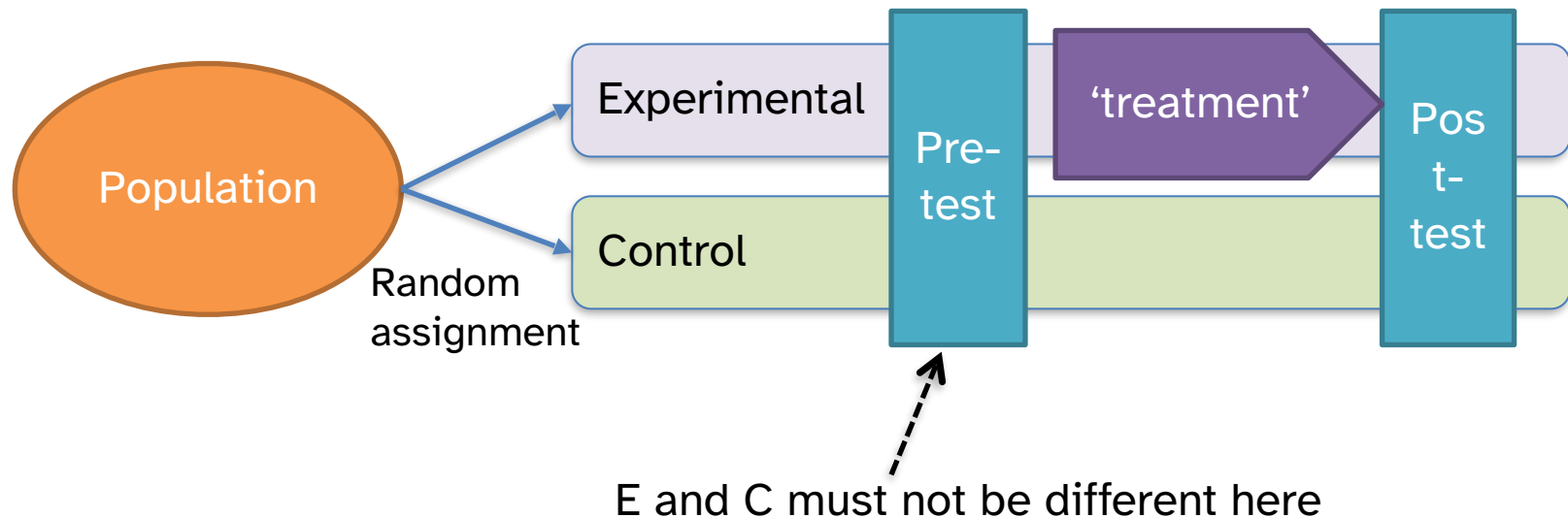
		Factor 1	
		Condition X	Condition Y
Factor 2	Condition P		
	Condition Q		

Example: Factor 1 (X = explicit corrections; Y = no corrections)
 Factor 2 (P = morphology; Q = syntax)

Important design concepts

Control group

A group that functions as the comparison for the experimental group in a between-subjects design



Participant recruitment

How many participants (N)?

Rule of thumb

- Look up similar studies in the literature
- Check the N size of those studies and try to test as many

Participant recruitment

How many participants (N)?

Power analysis

- Think about the **smallest effect size** that you would consider meaningful.
- Calculate **required sample size** to obtain the required statistical power (**minimum 80% power** recommended) to detect that effect size. (This can be done with the `simr` package)

Participant recruitment

Matching and screening

- Matching: Ensuring that the participants between groups are similar in all respects (e.g., age, sex) except the critical condition
- Screening: Ensuring that the participants meet certain conditions that can only be checked by some measurement (e.g., proficiency test)
- Both constrain recruitment

Participant recruitment

Compensation

- Monetary: £5/30 min for a simple experiment is standard
- Strategies to reduce monetary expenditure
 - Prize draw: e.g., Winner gets a gift voucher
 - Hand-made rewards: e.g., baked goods
 - Participation swap
- Use of mailing lists. See [PPLS policy](#). Include ethics permission statement.

Ethics: General principles

1. **Voluntary participation:** Participants must not be coerced and must be allowed to withdraw from studies.
2. **Informed consent:** Participants must give their consent to participate after being fully informed about the procedures, risks, and how their data might be used.
3. **Protection:** Participants should not be exposed to physical or mental harm.
4. **Anonymity:** The identity of the participants must not be revealed without their consent.
5. **No deception:** Do not misrepresent or withhold information unless it is absolutely necessary for the study (and even so, check with supervisor/committee).

Ethics: LEL procedure

1. Complete the *Research Ethics Training Course* on Learn.
2. Prepare components for your ethics application
 1. Answers to ethics questions
 2. Information sheet and consent form (templates available)
 3. Recruitment texts, if relevant
3. Meet with your supervisor to submit the application.
4. Wait for a reviewer to get back to you. It can take up to 10 working days to hear back.
5. Do not start data collection until you receive approval.

Details at [LEL Ethics page](#). (Or [Psychology Ethics page](#))

Ethics: Vulnerable groups

Vulnerable groups

- Children under the age of 16 (without adult supervision)
- Adults in care, prison etc.

Basic Disclosure

- A one-time background-check for relevant convictions in the UK. Typically asked if testing in schools.
- Can apply for it [online](#).

Protection of Vulnerable Group (PVG) scheme

- For longer-term involvement (caring, teaching etc.)
- Contact College PG Office for application (hsspg@ed.ac.uk)

Data management

What counts as data?

- Recordings
- Transcripts
- Questionnaires
- Numerical or textual data of responses in tests, experiments etc.
- Stimuli
- Consent forms

Data management

Guiding principles

- 1. Avoid recollecting data:** Mark the presence of your data so that someone else doesn't have to collect the same data.
- 2. Don't lose data:** Make sure they are kept properly and note where they are.
- 3. Keep data complete:** Look after the meta-data so that the data can be useful.
- 4. Identify ownership:** Establish who owns the data and who is responsible for looking after them (data steward).
- 5. Control data lifecycle:** Decide what is going to happen to the data (i.e., access, retention/destruction).

Data management

Guiding principles (cont.)

6. Store data securely:

- Anonymised/non-sensitive data: Store on password-protected personal computer or secure web-based server. Regular back-ups
- Non-anonymous/sensitive data: Store on encrypted university storage profile (automatic backup)

Summary

- In choosing a method suitable for your research question/ hypotheses, consider three aspects of methodology: design type, grouping and data type.
- In designing a study, maximise control and power by giving thoughts to the use of between- vs within-subjects design, control group, and factorial design.
- Determine the N size by consulting previous studies.
- Follow the principles of ethics and data management for research.