EMIA6500R · CMAA5022 · CSM160021

Social Media for Creatives
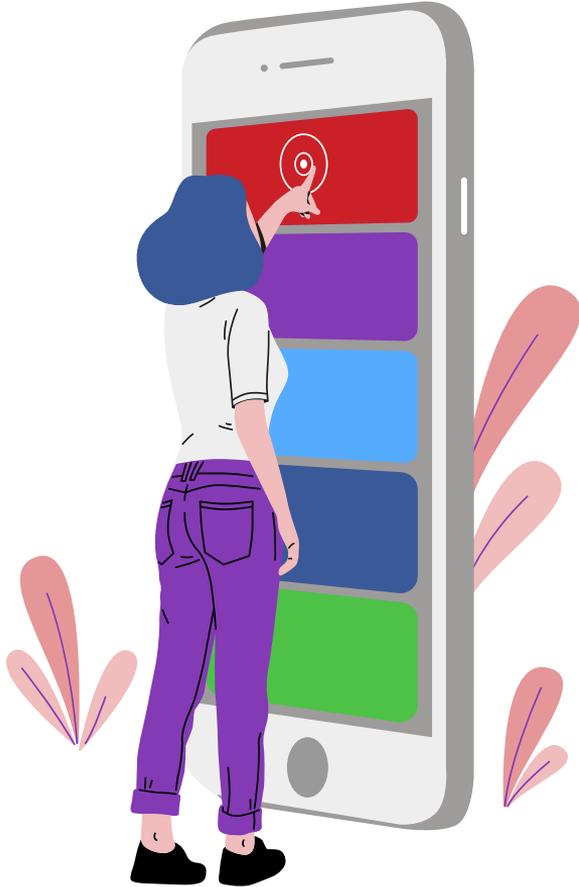
Lecture 07

Social Research Design

Professor Pan Hui
FRI 15:00-18:00 (HKT); 09:00-12:00 (EET)

Originally from Dr. Reza H. Mogavi

Today's Outline

01 **Lecturing**
Lecture 7: Social Research Design

02 **TA Session**
Jia's Tutorial

03 **Paper Reading Group Presentation**
Group 3

04 **VR Discussion & Participation**
Discuss with your groupmates in MetaClassroom

Learning Objectives

- Formulate clear and focused research questions

- Understand the need for quantitative research

- Learn the basic terminologies

- Differentiate between various types of research designs

What is Research?

- Research is a systematic process of collecting and analyzing data to enhance understanding of a phenomenon.
  ✅ True / ❌ False

- Research always involves experiments and hypothesis testing.
  ✅ True / ❌ False

- Research is only valid if it follows the scientific method.
  ✅ True / ❌ False

Scientific Method

The scientific method is primarily used in natural and social sciences (hypothesis, experimentation, observation).

The Scientific Method as an Ongoing Process
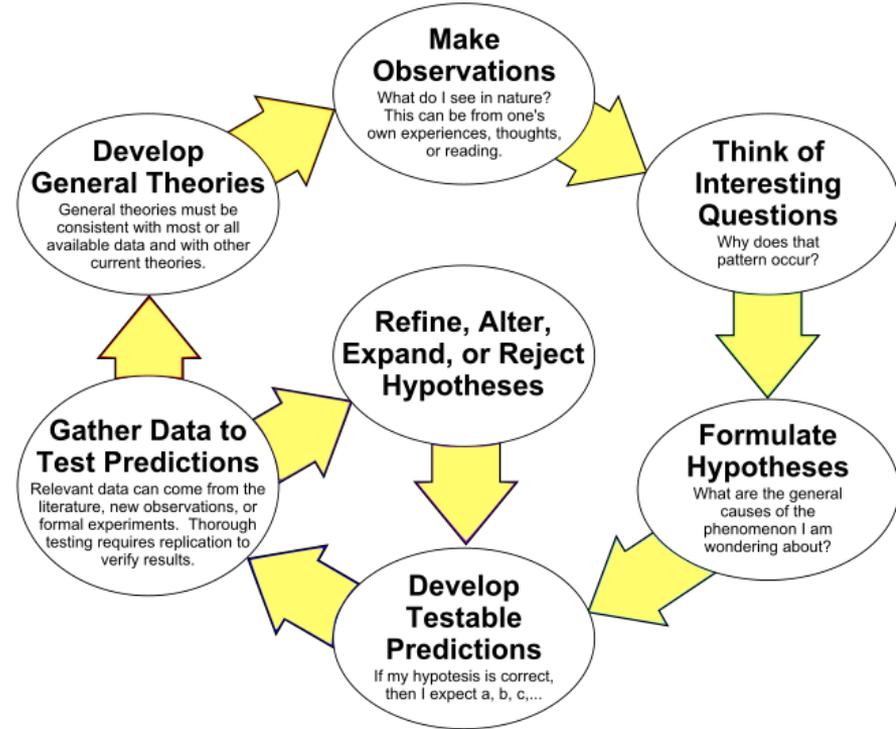
Research Question

Good research questions are:

- **Well-Motivated** – Clearly justified and relevant.
- **Clear** – Focused and well-defined.
- **Researchable** – Can be answered with data and analysis.
- **Specific** – Narrow enough to provide meaningful insights.
- **Open-Ended** – Avoids simple Yes/No answers
- **Ethical** – Follows established ethical guidelines, considering risks, fairness, and societal impact.
- **Useful** – Contributes to knowledge, decision-making, or problem-solving.

Hypothesis

- A hypothesis is a testable (falsifiable) statement that predicts a relationship between variables.

- Key features of a hypothesis:

  - **Clear & Specific**: Defines the expected relationship.

  - **Based on Theory or Prior Research**: Grounded in existing knowledge.

  - Provide **direction** for data collection and analysis.

Population vs. Sample

- A population is the complete set of all items of interest. Population size, N, can be very large or even infinite.

- A sample is an observed subset (or portion) of a population with sample size given by n.

- We hope the sample can represent the population, since our decision is made on the population.

Simple Random Sampling

It is a sampling scheme in that:

- Each member of the population has **the same probability** of being selected

- The selection of one member is **independent** of the selection of any other member

- Every possible sample of a given size, n, has the same probability of selection

Other Sampling Methods

Stratified Random Sampling:

- Divide the population into subgroups (strata) based on characteristics.
- Randomly select participants from each subgroup.
- Ensures proportional representation.

Systematic Sampling:

- Select every Nth person from a list after a random starting point.
- Simple and evenly distributed but may introduce hidden bias.

Other Sampling Methods

Cluster Sampling:

- Divide the population into clusters (e.g., schools, cities) and randomly select entire clusters.
- Cost-effective for large populations but may not be fully representative.

Convenience Sampling:

- Select participants who are easiest to reach.
- Quick and inexpensive but highly biased.

Other Sampling Methods

Purposive (Judgmental) Sampling:

- Choose participants based on **specific characteristics or expertise**.
- Useful for specialized studies but not random.

Snowball Sampling:

- Recruit participants through referrals from existing participants.
- Good for hard-to-reach populations but may lead to **similar perspectives**.

Mini Class Activity

A research team is conducting a study on how different age groups interact with an AI-powered healthcare assistant. They want to ensure that their findings reflect the experiences of teenagers, adults, and elderly users equally.

- Which sampling method would be the most appropriate? Why?

(A) Simple Random Sampling (B) Stratified Sampling

(C) Cluster Sampling (D) Convenience Sampling

Variables

Definition: A characteristic or condition that varies among individuals.

- Independent Variable (IV): The variable that is manipulated to observe its effect.

- Dependent Variable (DV): The variable that is measured to assess changes caused by the IV.

Defining Variables

Some variables are naturally easy to define

Examples: Time, height, weight, temperature.

These variables have a clear, universal measurement system (e.g., seconds, meters, kilograms, degrees)

Conceptual vs. Operational

Not every variable is easy to define!

Conceptual Definition: The abstract, theoretical meaning of a variable.

- Example: Stress is a psychological response to pressure.

Operational Definition: The specific, measurable way a variable is defined for a study.

- Example: Stress is measured by cortisol levels or self-reported anxiety scores.

Mini Class Activity

Develop a conceptual and operational definition for AI transparency

The Most Common Research Methods

Descriptive Methods:

- May lead to claims regarding the frequency of some behavior/observations

Correlational Methods:

- May lead to claims regarding the association between two/more variables
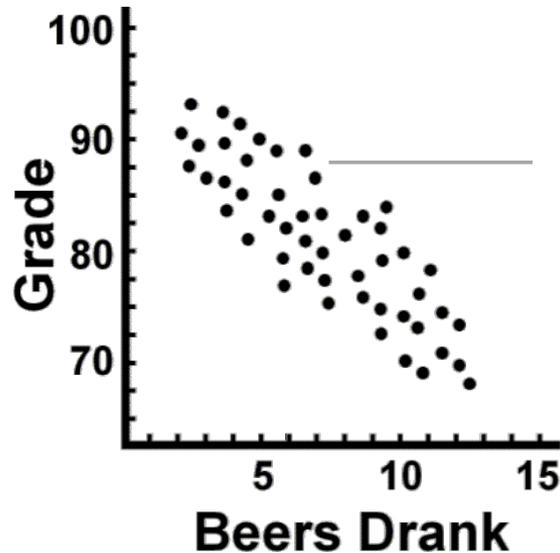
Experimental Methods:

- May lead to claims regarding the causal relationship between two/more variables

The Story of Clever Hans

https://www.youtube.com/watch?v=hAJlAuEo7Ac&t=12s&ab_channel=AnimalWised

Correlational Methods

Involve examining how variables are related (without manipulating any of the variables)
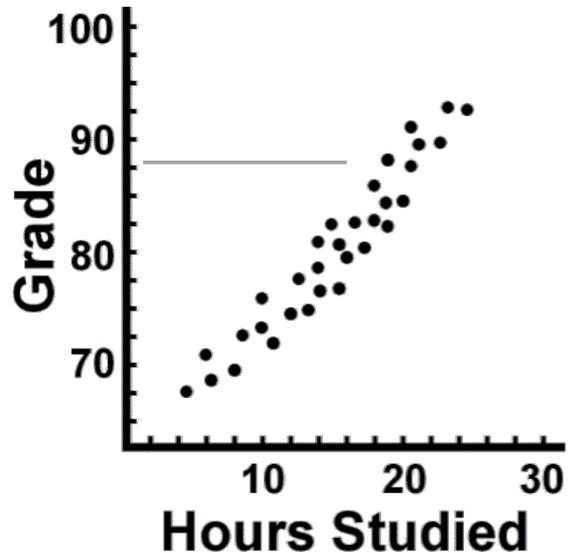
Did Ice Cream Cause Polio?

https://www.youtube.com/watch?v=lbODqslc4Tg&ab_channel=FreakonomicsRadioNetwork

Correlational Methods

Correlation is not causation.

Two main issues:
- Directionality problem
- Third variable problem
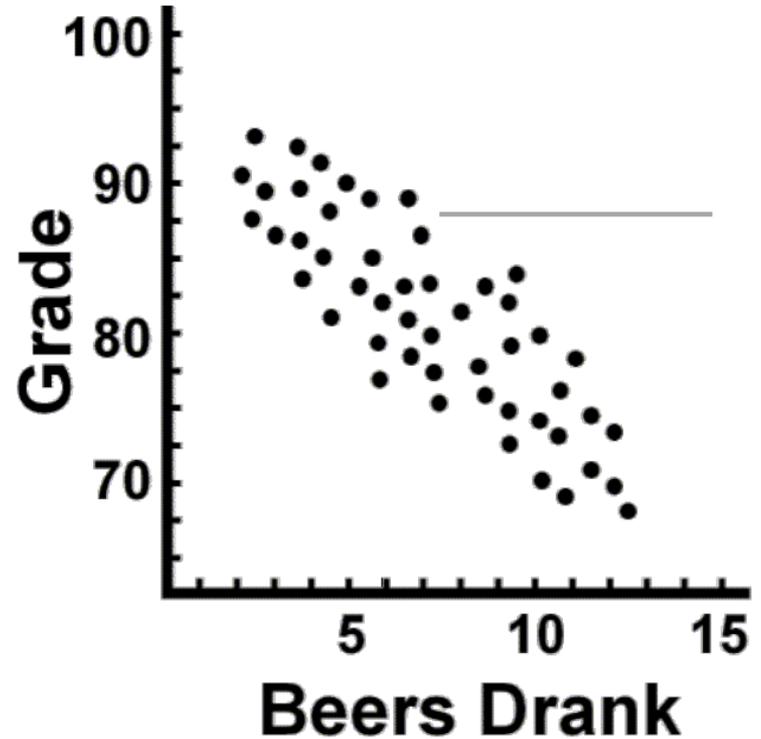
Experimental Methods

Between-Group Design

- Different groups of participants experience different experimental conditions.

Within-Group Design

- The same participants experience all experimental conditions.

Split-Plot Design

- A mix of between-group and within-group designs.

Between-Group Design

Example: Group A uses LLM 1, Group B uses LLM 2, and their performance is compared

Pros and Cons

- ==Minimizes learning effects== since participants don't experience other conditions.

- ==Reduces fatigue and frustration== due to shorter sessions.

- Requires a ==larger sample size== to account for individual differences.

Within-Group Design

Example: Users first try LLM 1, then LLM 2, and their experiences are compared

Pros and Cons

- Controls for **individual variability** among participants.
- Requires a **smaller sample size** compared to between-group designs.
- May introduce learning effects and fatigue over time.

Split-Plot Design

Example:

Between-Group Factor:

Users are divided into two groups:

- Group A uses LLM 1
- Group B uses LLM 2

Within-Group Factor:

Each user is tasked to ask three types of question

- Factual
- Reasoning
- Creative

Split-Plot Design

Pros and Cons

- Suitable for complex studies with multiple independent variables.

- It is commonly used when some factors are easier to manipulate between subjects, while others must be tested within subjects.

- Allows examination of interaction effects between variables.

- Requires careful experimental control to manage complexity and accuracy.

Confounding & Controlling Variables

- Anything that may unintentionally vary along with the independent variable

- Confounds limit our ability to make causal claims

- Proper control of variables is vital for validity

Counterbalancing

- Counterbalancing is a research technique used to control order effects in repeated-measures experiments, where participants experience multiple conditions.

- It ensures that the order of conditions does not bias the results.

- Prevents **learning effects** (improvement from practice).

- Minimizes **carryover effects** (previous condition affecting the next).

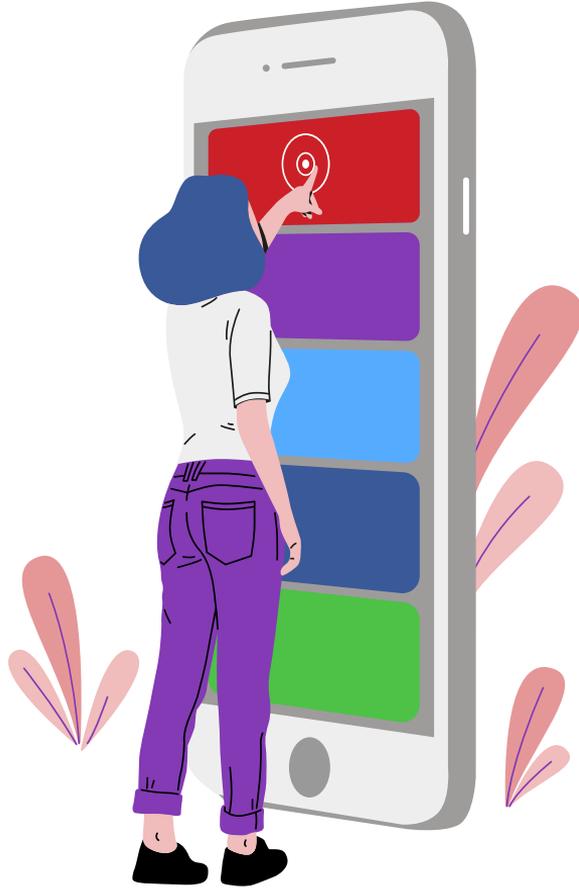# Resources to Explore this Topic Further
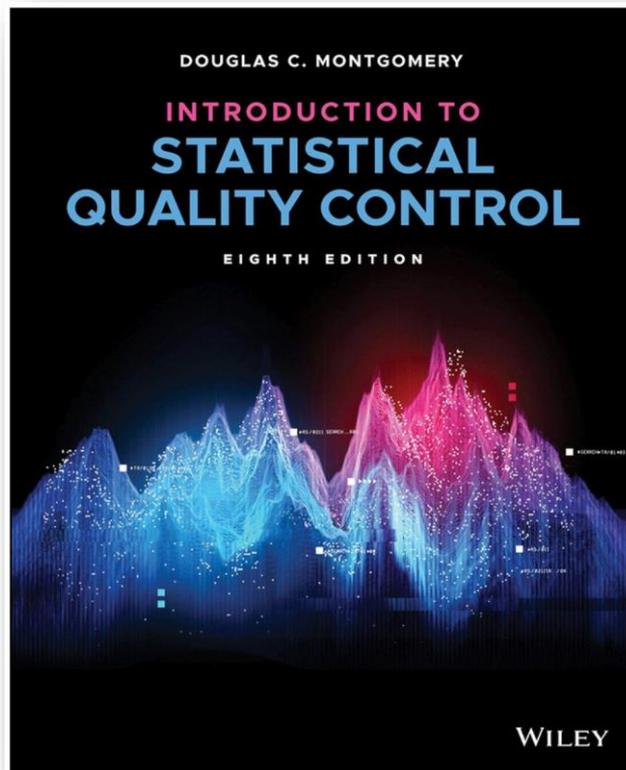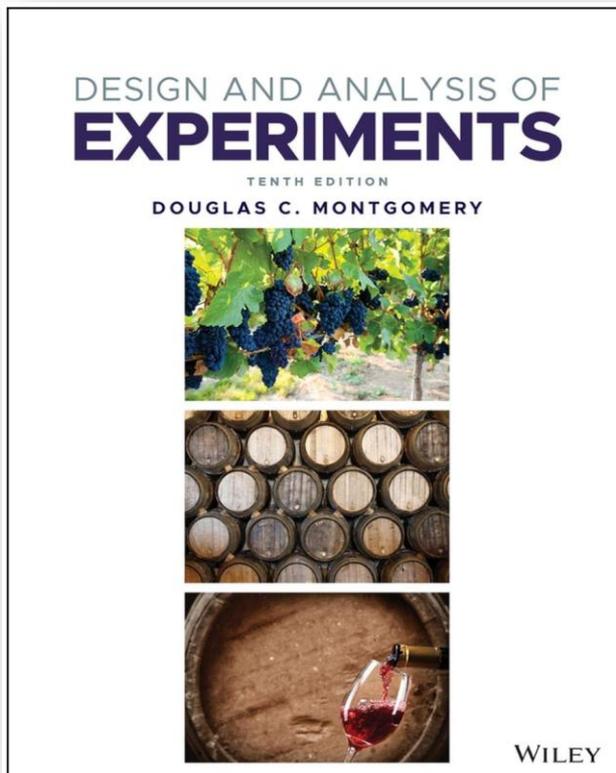
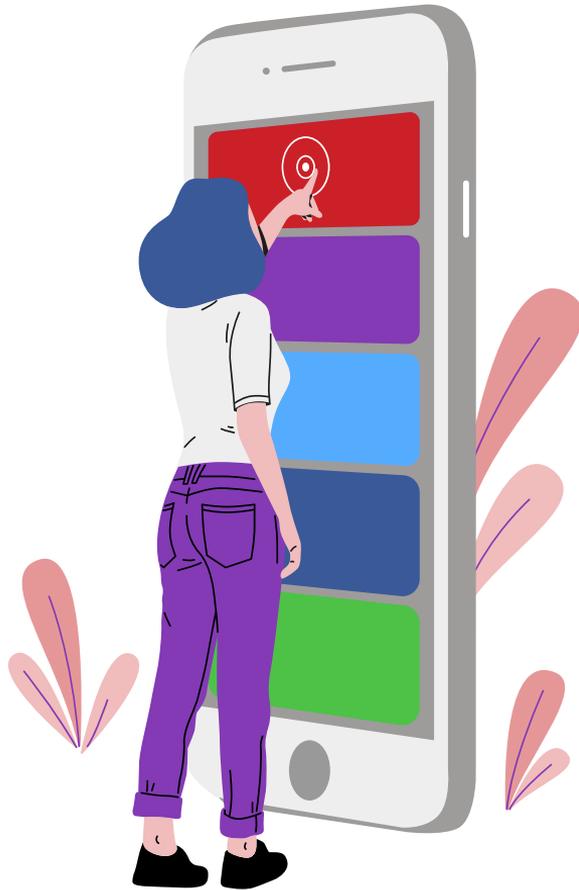# Today's Outline
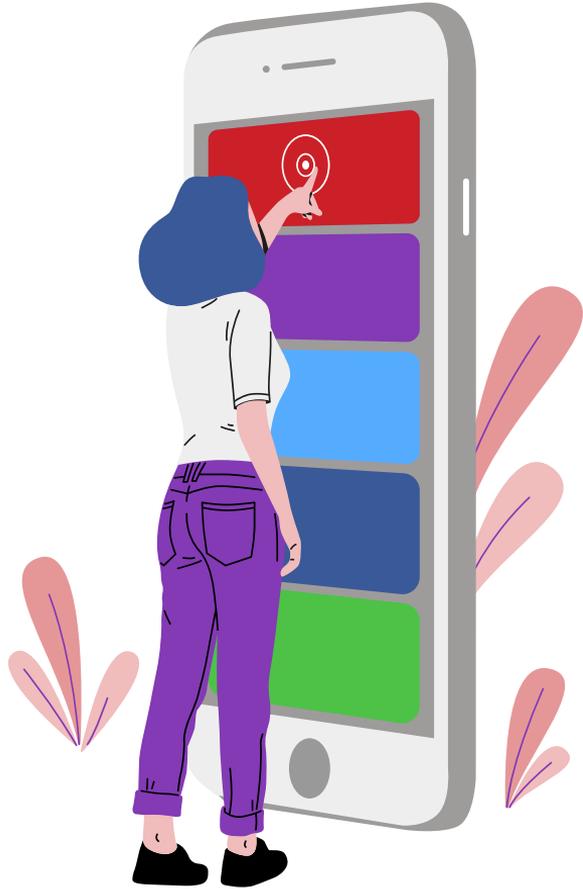
TA Session
Jia's Tutorial

Today's Outline

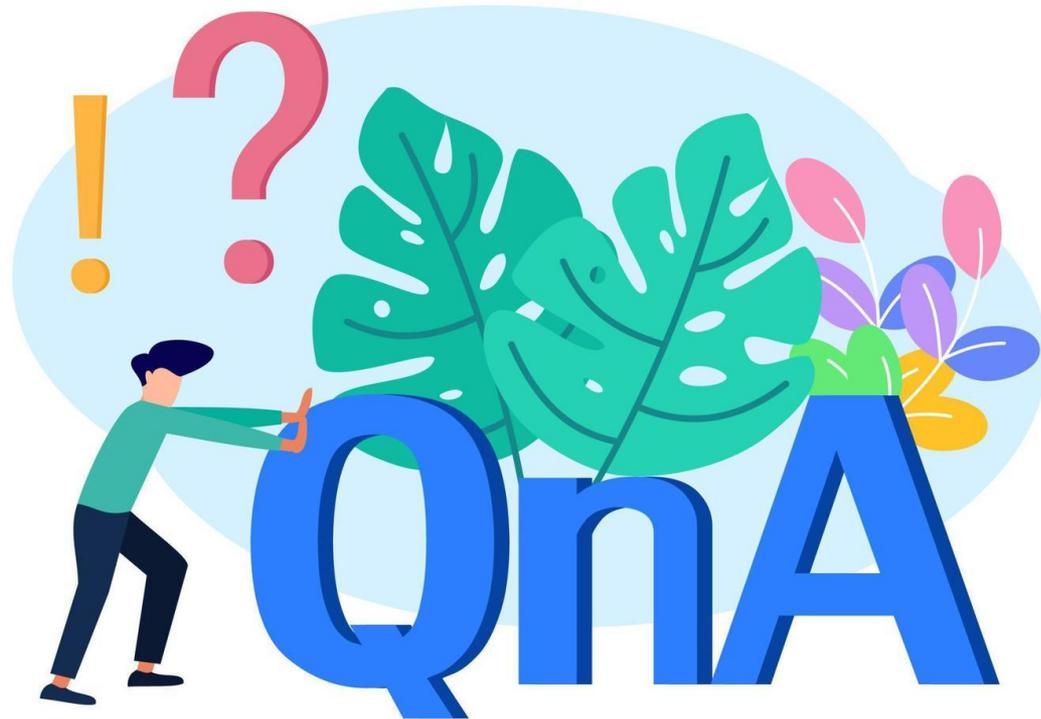**Paper Reading Group Presentation**

Group 3: 15 minutes presentation

Today's Outline

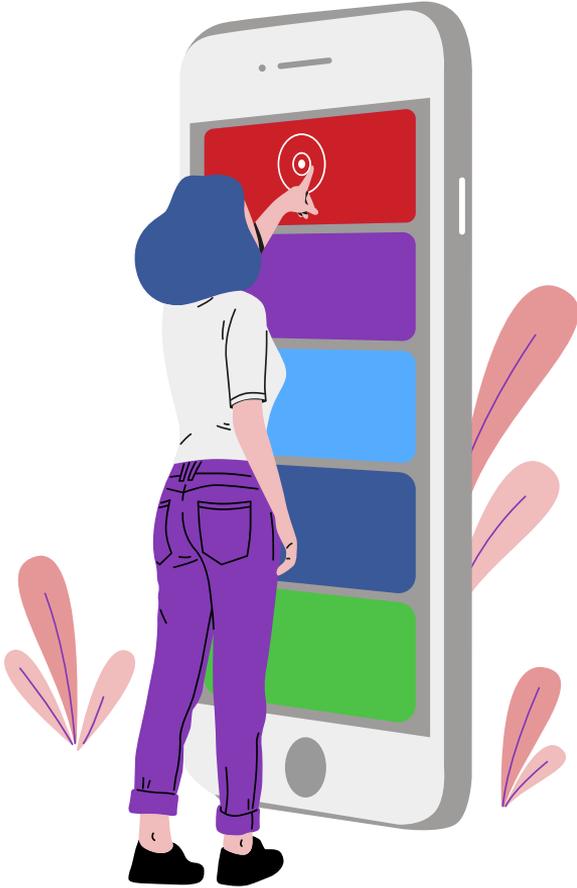**Paper Reading Group Presentation**
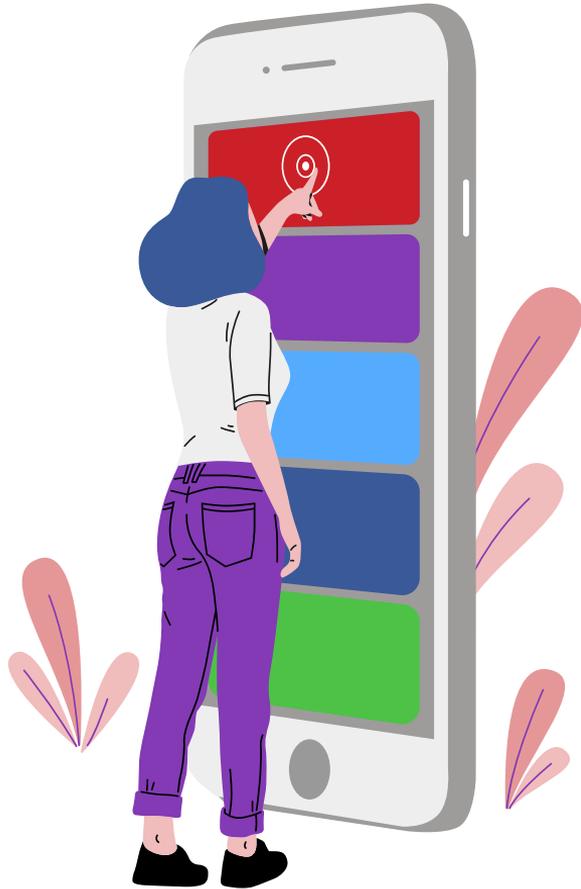Group 3: QnA (5 minutes)

Today's Outline

VR Discussion & Participation

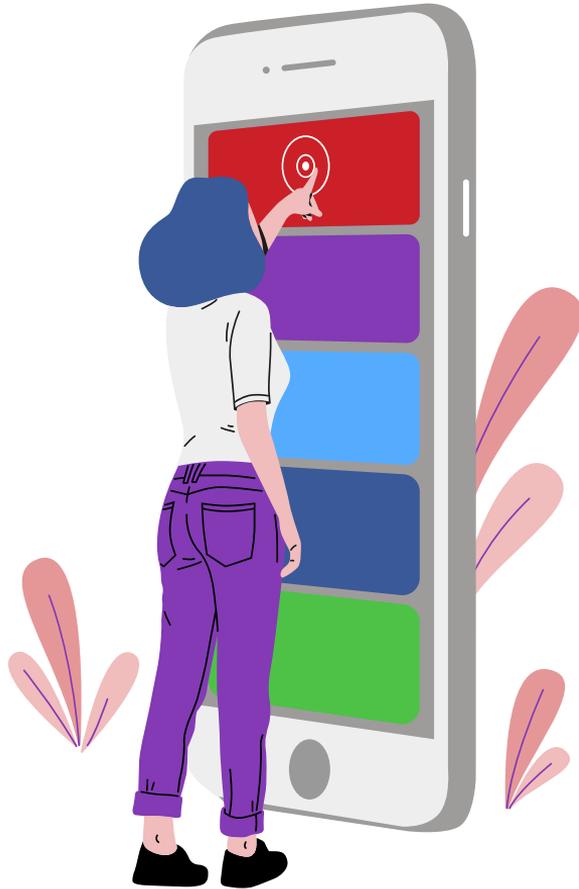Participation

Today's Outline

VR Discussion & Participation

Discuss with your groupmates in MetaClassroom

Task 1: Individual

How would you study this phenomenon?	Member 1	Member 2	Member 3	Member 4
What methods fit (surveys, interviews, content analysis, experiments, network analysis)?	Member 1	Member 2	Member 3	Member 4
What data is available?	Member 1	Member 2	Member 3	Member 4
What questions could you answer?	Member 1	Member 2	Member 3	Member 4

Today's Outline

VR Discussion & Participation
Discuss with your groupmates in MetaClassroom

Task 2: VR Group
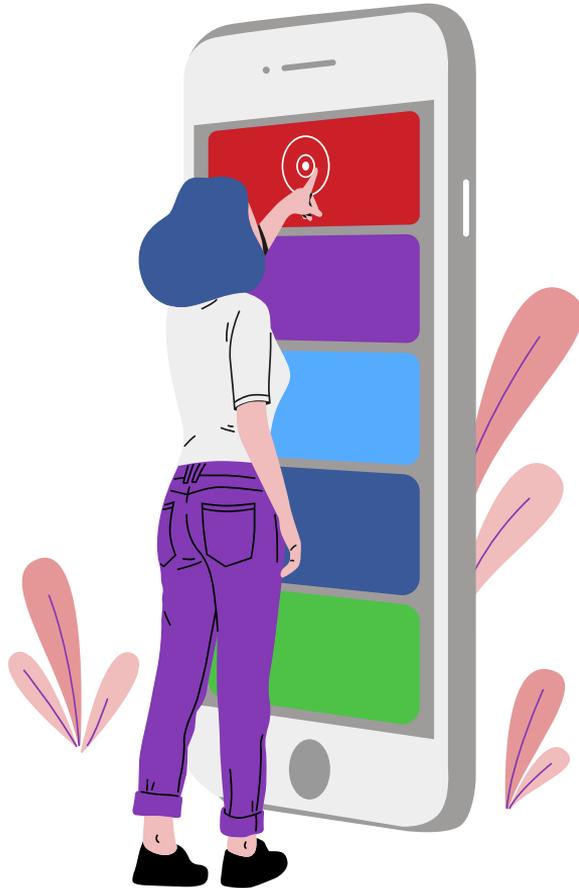
Do we agree on the best methods?

What would quantitative vs qualitative approaches reveal?

What's hard to study about this phenomenon?

What ethical issues might arise?

Identify one strong research question, one methodological disagreement.

Today's Outline

VR Discussion & Participation
Discuss with your groupmates in MetaClassroom

Task 3: After VR

proposed research question, methods that would work, data sources available, limitations and ethics, connections to previous observations.	Answer	Answer	Answer	Answer
	Answer	Answer	Answer	Answer
Individual reflection: What would you most want to know about this phenomenon? Why is it hard to answer?	Member 1	Member 2	Member 3	Member 4
	Member 1	Member 2	Member 3	Member 4