

# VU LEARNING HUB

[vulearninghub.xyz](https://vulearninghub.xyz)

**HIGHLIGHTED HANDOUT**

## **CS719**

Research Methodologies

**CS719 by VU Learning HUB**

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

VU STUDY HANDOUT

# CS719

Research Methodologies

This handout builds end-to-end research competence for computing and social-science researchers: formulating problems, reviewing literature, choosing designs, sampling, collecting and analyzing data, and writing publishable papers ethically.

15 lecture-wise topics | key terms | exam pointers



# Table of Contents

## UNIT 1: FOUNDATIONS OF RESEARCH

Lecture 1: What is Research?

Lecture 2: Formulating the Research Problem

Lecture 3: Literature Review

## UNIT 2: RESEARCH DESIGNS

Lecture 4: Quantitative Research Designs

Lecture 5: Qualitative Research Designs

Lecture 6: Mixed Methods and Survey Research

## UNIT 3: SAMPLING AND DATA COLLECTION

Lecture 7: Sampling Techniques

Lecture 8: Measurement: Validity and Reliability

Lecture 9: Data Collection Methods

## UNIT 4: DATA ANALYSIS

Lecture 10: Quantitative Data Analysis

Lecture 11: Qualitative Data Analysis

## UNIT 5: REPORTING AND ETHICS

Lecture 12: Writing the Research Report/Thesis

Lecture 13: Referencing and Avoiding Plagiarism

Lecture 14: Research Ethics

Lecture 15: Publishing and Research Impact

# UNIT 1: FOUNDATIONS OF RESEARCH

## Lecture 1: What is Research?

### Learning Objectives

- Define research and its characteristics
- Distinguish research types
- Explain the research process

### Key Terms & Definitions

#### Research

Systematic, controlled, empirical and critical investigation of a problem to discover new knowledge.

#### Basic vs Applied research

Knowledge for its own sake vs knowledge solving a practical problem.

#### Hypothesis

A testable tentative proposition about relationships between variables.

Research is **systematic inquiry**: it follows an explicit method, controls bias, grounds claims in **evidence**, and opens itself to **scrutiny**. Its hallmarks: purposiveness, rigor, testability, replicability, precision, objectivity, and parsimony. Casual Googling is not research; a documented, reproducible investigation is.

**Types**: basic/pure (extends theory), applied (solves real problems), **action research** (practitioner improves own practice iteratively), evaluation research, and **R&D**. In computing, applied/experimental research dominates: build a system, measure it against baselines. **Quantitative** research counts and models; **qualitative** explores meanings; **mixed methods** triangulate.

The **research process**: identify problem -> literature review -> formulate objectives/hypotheses -> design -> collect data -> analyze -> interpret -> report. It is **iterative**: findings send you back to refine questions. Every choice must be **justified** in the thesis - "why this method" matters as much as results.

- **Research gap**: the unanswered question your study fills - the heart of any proposal.
- **Variables**: independent (manipulated), dependent (measured), control/extraneous.
- **Deductive** (theory -> test) vs **inductive** (observations -> theory).
- **Paradigms**: positivism (objective measurement), interpretivism (meaning), pragmatism (what works).

#### VERY IMPORTANT — Exam Pointer

Very important: define research with its characteristics, and distinguish basic vs applied with computing examples.

### Section Summary

Research is disciplined, evidence-based inquiry aimed at a clearly identified gap.

## Lecture 2: Formulating the Research Problem

### Learning Objectives

- Identify sources of research problems
- Write good research questions
- Evaluate problem statements

### Key Terms & Definitions

#### Problem statement

Clear articulation of the gap, its significance, and the study's intent.

#### Research questions

Specific interrogatives the study will answer.

#### Delimitations

Boundaries the researcher deliberately sets (scope).

Problems come from **literature gaps** (contradictions, untested contexts), **practice** (industry pain points), **theory** (unverified predictions), and **prior findings** (a paper's "future work"). In CS: "existing load balancers ignore energy cost in edge data centers" is a problem; "I want to study networks" is not.

Good **research questions** are focused, researchable, and significant: "How does algorithm X affect latency under bursty IoT traffic compared to Y?" Criteria: **FINER** (Feasible, Interesting, Novel, Ethical, Relevant). Each question maps to objectives and later to findings - maintain this **golden thread**.

The **problem statement** has three moves: the **context** (what is known), the **gap** (what is missing/wrong), the **significance** (why filling it matters, for whom). State **objectives** (SMART), **scope/delimitations**, and operational **definitions** of key terms.

- **FINER criteria** for evaluating a research question.
- **Golden thread**: questions -> objectives -> method -> findings -> conclusions must align.
- **Scope**: say what the study will NOT cover - examiners probe this.
- **Avoid**: questions answerable by yes/no, or by a quick search.

#### Example

Weak: "Study of e-learning." Strong: "To what extent does adaptive quiz sequencing improve retention among VU distance learners compared to fixed sequencing?"

#### VERY IMPORTANT — Exam Pointer

Exam pointer: convert a vague topic into a FINER research question with objectives.

### Section Summary

A sharp problem statement and FINER questions are the foundation everything else stands on.

## Lecture 3: Literature Review

### Learning Objectives

- Explain the purposes of literature review
- Search and organize literature
- Write a critical synthesis

### Key Terms & Definitions

#### Literature review

Critical synthesis of existing knowledge identifying gaps and positioning the study.

#### Systematic review

Protocol-driven, reproducible review (e.g. PRISMA) minimizing selection bias.

#### Citation databases

IEEE Xplore, ACM DL, Scopus, Web of Science, Google Scholar.

The review **justifies** the study: it maps what is known, exposes the **gap**, positions your contribution, and informs method choices. A good review is **critical and synthetic** - organized by themes/approaches, comparing and evaluating - not an annotated bibliography summarizing papers one by one.

**Searching:** start with seed papers, then **forward/backward snowballing** (citations), database queries with Boolean operators, and alerts. Track with a reference manager (**Zotero/Mendeley**). Assess quality: venue reputation, citations, method rigor. For a thesis, a **systematic** protocol (inclusion/exclusion criteria, PRISMA flow) adds credibility.

**Writing:** thematic structure (chronological only when the evolution matters), **synthesis matrices** (papers x themes), critical voice ("However, these studies assume...; none addresses..."), ending by **stating the gap** your study fills. Every claim cited; every citation earned.

- **Synthesize, don't list:** group by theme, compare findings, note contradictions.
- **Snowballing** from 2-3 seed papers beats keyword searching alone.
- **PRISMA:** identification -> screening -> eligibility -> included.
- **Theoretical framework:** the lens (e.g. TAM) guiding your study, derived from the review.

#### VERY IMPORTANT — Exam Pointer

Very important: distinguish a critical synthesis from an annotated bibliography.

### Section Summary

The literature review earns your study's right to exist by critically mapping the gap.

## UNIT 2: RESEARCH DESIGNS

### Lecture 4: Quantitative Research Designs

## Learning Objectives

- Classify quantitative designs
- Explain experimental design
- Discuss validity threats

## Key Terms & Definitions

### Experimental design

Manipulates an independent variable with random assignment and control.

### Quasi-experimental

Lacks random assignment; uses intact groups.

### Internal validity

Confidence that the treatment caused the observed effect.

**Experimental** (true experiment): random assignment + control group + manipulation - the gold standard for **causality**. **Quasi-experimental**: no random assignment (e.g. two existing classes) - weaker causal claims. **Non-experimental**: descriptive surveys, correlational studies, ex post facto - no manipulation.

In computing, experiments are **benchmark studies**: algorithm variants (treatments) on datasets (subjects), measuring latency/accuracy with **repeated runs**, controlling hardware/caching confounds, using **statistical tests** (t-test, ANOVA) rather than eyeballing means. **Ablation studies** isolate each component's contribution.

**Validity**: internal (confounds, history, maturation, testing effects), external (generalizability - ecological validity of lab benchmarks), construct (does the metric measure the concept?), conclusion (statistical power). **Reliability** (consistency) is necessary but not sufficient for validity.

- **Randomization** is what separates true experiments from quasi.
- **Control confounds**: same hardware, warm-up runs, fixed seeds.
- **Replication** (multiple runs) quantifies variance - report std dev / confidence intervals.
- **Threats checklist** belongs in every thesis methodology.

### Example

Comparing two schedulers: same workload traces, 30 runs each, random order, t-test on mean job completion time - a true experimental design in systems research.

### VERY IMPORTANT — Exam Pointer

Exam pointer: name four threats to internal validity and one control for each.

## Section Summary

Quantitative designs trade off control against realism; validity analysis shows you understand the trade.

## Lecture 5: Qualitative Research Designs

### Learning Objectives

- Describe major qualitative traditions
- Explain purposive sampling
- Discuss trustworthiness

## Key Terms & Definitions

### Case study

In-depth investigation of a bounded system (a team, a project, an organization).

### Grounded theory

Builds theory inductively from data via coding.

### Thematic analysis

Identifies patterns of meaning across qualitative data (Braun & Clarke).

**Traditions:** case study (the "how/why" of a real context - common in software engineering research), **ethnography** (immersion in a culture, e.g. open-source communities), **phenomenology** (lived experience), **grounded theory** (theory emerging from data), **action research** (intervene + study). Choose by question: "how" questions suit cases; "what is the experience" suits phenomenology.

**Data:** semi-structured **interviews** (flexible protocol), focus groups, observation, documents/artifacts. **Purposive sampling:** select information-rich cases (maximum variation, snowball) - not random. **Saturation** (no new themes emerging) decides sample size, often 12-30 interviews.

**Analysis:** transcribe -> **open coding** (label meaning units) -> axial/selective coding -> themes (thematic analysis) or a theory (grounded theory). **Trustworthiness:** credibility (member checking, triangulation), transferability (thick description), dependability (audit trail), confirmability (reflexivity). Software: NVivo, ATLAS.ti.

- **"How/why" questions -> qualitative;** "how much/how many" -> quantitative.
- **Saturation,** not a formula, justifies qualitative sample size.
- **Triangulation:** interviews + documents + observation converging.
- **Reflexivity:** document your own influence on the data.

### VERY IMPORTANT — Exam Pointer

Very important: explain credibility, transferability, dependability, confirmability in one line each.

## Section Summary

Qualitative designs explore meaning in context; rigor comes from transparency and triangulation.

## Lecture 6: Mixed Methods and Survey Research

### Learning Objectives

- Explain mixed-method designs
- Design valid questionnaires
- Avoid survey pitfalls

## Key Terms & Definitions

### Triangulation design

Quantitative + qualitative concurrently to corroborate findings.

### Explanatory sequential

Quantitative first, then qualitative to explain results.

### Likert scale

Ordered agreement scale (typically 5- or 7-point) for attitude measurement.

**Mixed methods:** triangulation (concurrent, corroborate), **explanatory sequential** (survey then interviews explaining surprises), **exploratory sequential** (interviews then survey to generalize). Pragmatist paradigm: use what answers the question. Integration must happen - in analysis or interpretation - not two separate studies stapled together.

**Questionnaire design:** one concept per item, simple unambiguous wording, no double-barreled or leading questions, balanced scales, pilot test (n=10-20) with cognitive debriefing. **Likert** items need labeled points; validate constructs with **factor analysis** and report **Cronbach's alpha** ( $\geq 0.7$ ) for reliability.

**Survey errors:** coverage (bad frame), sampling, nonresponse (follow-ups, incentives), measurement (poor items). Online surveys risk self-selection; report response rates honestly. In CS: developer surveys (Stack Overflow style) follow identical principles.

- **Pilot everything:** a 15-minute pilot saves a 300-response disaster.
- **Cronbach's alpha  $\geq 0.7$**  for scale reliability.
- **Avoid:** double-barreled ("fast and reliable?"), leading, and absolute ("always") items.
- **Mixed methods need integration** - a joint display or merged interpretation.

### Example

Explanatory sequential: survey finds 68% of developers distrust AI code review; follow-up interviews reveal the distrust centers on security-blind approvals - the qualitative explains the quantitative.

### VERY IMPORTANT — Exam Pointer

Exam pointer: name the three mixed-method designs and when each fits.

## Section Summary

Surveys quantify attitudes at scale; mixed methods combine numbers with the reasons behind them.

# UNIT 3: SAMPLING AND DATA COLLECTION

## Lecture 7: Sampling Techniques

### Learning Objectives

- Distinguish probability and non-probability sampling

- Compute sample size logic
- Avoid sampling bias

## Key Terms & Definitions

### Probability sampling

Every element has a known, non-zero chance of selection.

### Stratified sampling

Population divided into strata; random sample within each.

### Sampling frame

The list from which the sample is drawn.

**Probability:** simple random, **stratified** (by department/year - guarantees representation), **cluster** (sample whole groups, e.g. classes - cheap for dispersed populations), systematic (every kth). Only probability samples support **statistical generalization** with measurable sampling error.

**Non-probability:** convenience, purposive, snowball, quota. Legitimate for qualitative/exploratory work and when no frame exists - but findings **do not generalize** statistically; say so explicitly. Most student surveys are convenience samples: acknowledge the limitation.

**Sample size:** for estimation, n depends on desired margin of error, confidence level, and population variance (Cochran's formula); for experiments, **power analysis** (G\*Power: effect size, alpha 0.05, power 0.80). Rules of thumb (n=30, 10% of population) are weak - justify with power or saturation.

- **Stratified > simple random** when subgroups must be represented.
- **Power analysis** justifies experimental sample sizes.
- **Sampling frame quality** bounds everything - a bad list means a bad sample.
- **Report:** technique, frame, size justification, response rate.

### VERY IMPORTANT — Exam Pointer

Very important: contrast probability vs non-probability sampling and state what each permits you to claim.

## Section Summary

Sampling determines what you may generalize; match the technique to the claim.

## Lecture 8: Measurement: Validity and Reliability

### Learning Objectives

- Explain levels of measurement
- Assess validity types
- Measure reliability

## Key Terms & Definitions

### Construct validity

The instrument truly measures the intended concept.

### Reliability

Consistency of measurement across time/items/raters.

### Operationalization

Turning a concept into measurable indicators.

**Levels:** nominal (categories), ordinal (ranked), interval (equal gaps, no true zero), ratio (true zero). The level dictates permissible statistics (means need interval+; chi-square for nominal). **Operationalization:** "usability" -> task completion time + SUS questionnaire score.

**Validity:** content (expert coverage), criterion (correlation with a gold standard), **construct** (convergent/discriminant evidence, factor analysis). A speed test that actually measures test-taking skill lacks construct validity.

**Reliability:** test-retest, **internal consistency** (Cronbach's alpha), inter-rater (**Cohen's kappa**). Unreliable measures attenuate real effects - pilot and refine. **Validity requires reliability**, but reliability alone is not validity.

- **Nominal < ordinal < interval < ratio:** know which statistics each allows.
- **Pilot + alpha + factor analysis:** the instrument validation trio.
- **Kappa** for agreement between human coders/raters.
- **Report psychometrics** for every adopted or adapted scale.

### Example

Adopting the System Usability Scale (SUS): 10 items, 5-point Likert, score 0-100; prior studies report alpha > 0.90 - cite this as reliability evidence.

### VERY IMPORTANT — Exam Pointer

Exam pointer: explain why reliability is necessary but not sufficient for validity.

## Section Summary

Measure the right thing (validity) consistently (reliability), and prove both.

## Lecture 9: Data Collection Methods

### Learning Objectives

- Select appropriate data collection methods
- Conduct effective interviews
- Use observation systematically

### Key Terms & Definitions

#### Semi-structured interview

Guided by a protocol but flexible to follow up.

### Structured observation

Predefined coding scheme for recording behavior.

### Secondary data

Existing datasets reused for new questions.

**Interviews:** structured (identical questions, comparable), **semi-structured** (protocol + probes - the qualitative workhorse), unstructured (conversational). Craft open questions, avoid leading, probe ("tell me more about..."), record with consent, transcribe verbatim.

**Observation:** participant vs non-participant; structured (checklists, e.g. counting help-desk requests) vs unstructured field notes. **Documents/artifacts:** commit logs, tickets, policies - unobtrusive and rich in software research. **Secondary data** (Kaggle, UCI, public APIs) is legitimate when the question fits - cite the source.

**Fieldwork discipline:** consent forms, anonymization at collection, secure storage, audit trail of decisions. Response quality beats quantity: ten deep interviews beat 200 careless questionnaires.

- **Match method to question:** behaviors -> observation; experiences -> interviews; traces -> artifacts.
- **Informed consent** always: purpose, voluntariness, anonymity, withdrawal right.
- **Triangulate** collection methods to offset each one's weakness.
- **Log everything:** dates, settings, non-responses, deviations.

### VERY IMPORTANT — Exam Pointer

Very important: list the ethical requirements of data collection (consent, anonymity, withdrawal).

### Section Summary

Collect data the question demands, ethically and traceably.

## UNIT 4: DATA ANALYSIS

### Lecture 10: Quantitative Data Analysis

#### Learning Objectives

- Choose appropriate statistical tests
- Interpret p-values and effect sizes
- Present results correctly

#### Key Terms & Definitions

##### p-value

Probability of data this extreme if the null hypothesis were true.

##### Effect size

Magnitude of the difference/relationship (Cohen's d, r, eta-squared).

### Type I / Type II error

False positive (alpha) / false negative (beta).

**Test selection:** compare two means -> t-test (paired if same subjects); 3+ groups -> ANOVA; categorical association -> chi-square; relationship -> correlation/regression; non-normal/small samples -> non-parametric (Mann-Whitney, Wilcoxon). Check **assumptions** (normality: Shapiro-Wilk; equal variance: Levene) before trusting the test.

**p < 0.05 is not the finding** - it only rejects the null. Report **effect sizes** (Cohen's d: 0.2 small, 0.5 medium, 0.8 large) and **confidence intervals**, which convey practical significance. **Multiple comparisons** inflate false positives - correct with Bonferroni/Holm.

**Presentation:** descriptives first (means, SDs, n), then inferential results in APA style:  $t(28) = 2.45, p = .021, d = 0.89$ . Tables for precision, figures for patterns. Never "prove" - you "found evidence for".

- **Assumptions first, test second.**
- **Effect size + CI** always alongside p-values.
- **Correct for multiple comparisons.**
- **Practical vs statistical significance:** huge n makes trivia "significant".

### Example

New cache policy: mean hit-rate 87.3% vs 82.1%,  $t(58)=3.12, p=.003, d=0.81$  (large) - practically meaningful, not just significant.

### VERY IMPORTANT — Exam Pointer

Very important: interpret a p-value correctly and explain why effect sizes are required.

### Section Summary

Statistics quantify evidence; effect sizes and intervals tell you whether it matters.

## Lecture 11: Qualitative Data Analysis

### Learning Objectives

- Code qualitative data
- Build themes systematically
- Use CAQDAS tools

### Key Terms & Definitions

#### Open coding

Labeling data segments with conceptual codes.

#### Axial coding

Relating codes into categories and relationships.

### Member checking

Returning interpretations to participants for validation.

**Thematic analysis** (Braun & Clarke, six phases): familiarize -> generate initial codes -> search themes -> review themes -> define/name -> write up. **Codes** are labels ("frustration with documentation"); **themes** are patterns answering the research question ("documentation as a barrier to onboarding").

**Grounded theory** goes further: constant comparison builds an explanatory **theory** (e.g. how teams adopt DevOps). Keep **memos** - analytic notes capturing your thinking; they become the theory's scaffolding. **Negative case analysis** strengthens claims.

**Tools** (NVivo, ATLAS.ti, MAXQDA) organize coding, not do thinking. **Inter-coder reliability** (Cohen's kappa on a subset) for team projects. Present themes with **evocative quotes**, counts used sparingly (qualitative != quantitative).

- **Code -> categorize -> theme**: the analytic ladder.
- **Memos** capture insight when it strikes - write them immediately.
- **Quotes evidence themes**; never present a theme without data.
- **Audit trail**: codebook versions, theme decisions, reflexive notes.

### VERY IMPORTANT — Exam Pointer

Exam pointer: outline the six phases of thematic analysis.

## Section Summary

Qualitative analysis turns transcripts into trustworthy themes through systematic coding.

# UNIT 5: REPORTING AND ETHICS

## Lecture 12: Writing the Research Report/Thesis

### Learning Objectives

- Structure a thesis/paper
- Write each chapter effectively
- Master academic style

### Key Terms & Definitions

#### IMRaD

Introduction, Methods, Results, and Discussion - the standard paper structure.

#### Abstract

150-250 word standalone summary: problem, method, key results, conclusion.

#### APA style

American Psychological Association formatting and citation standard.

**Thesis structure:** preliminary pages -> Chapter 1 Introduction (background, problem, questions, objectives, significance, scope, definitions) -> Chapter 2 Literature Review -> Chapter 3 Methodology (design, population, sampling, instruments, procedure, analysis, ethics) -> Chapter 4 Results -> Chapter 5 Discussion/Conclusion (summary, implications, limitations, future work) -> references -> appendices.

**Writing craft:** one idea per paragraph with a topic sentence; **hedge** claims appropriately ("suggests" not "proves"); past tense for your procedures/results, present for established knowledge; figures/tables numbered, titled, and **referenced in text**. Write the **abstract last**.

**Paper version (IMRaD):** conferences/journals want contribution up front - abstract, introduction ending in contributions, related work, method, evaluation, discussion, conclusion. Follow the venue's template exactly; page limits are hard.

- **Golden thread:** every chapter answers the research questions.
- **Methodology must enable replication** - enough detail to redo the study.
- **Discuss limitations honestly** - it strengthens, not weakens, credibility.
- **Abstract last**, and make it self-contained.

**VERY IMPORTANT — Exam Pointer** 

Very important: list thesis chapters and state the purpose of the methodology chapter.

## Section Summary

Structure carries the argument: IMRaD for papers, five chapters for theses, precision throughout.

## Lecture 13: Referencing and Avoiding Plagiarism

### Learning Objectives

- Apply APA citation correctly
- Distinguish plagiarism types
- Use Turnitin responsibly

### Key Terms & Definitions

#### Plagiarism

Presenting others' work/ideas as one's own, with or without intent.

#### Paraphrasing

Restating ideas in one's own words with citation.

#### Self-plagiarism

Reusing one's own published work without disclosure.

**APA 7th basics:** in-text (Author, Year); direct quotes need page numbers; reference list alphabetical with hanging indent: Author, A. (Year). Title. Source. **Cite** every idea, number, and distinctive phrase that is not yours

- including code and figures (adapted/redrawn with credit).

**Plagiarism spectrum:** copy-paste, mosaic/patchwriting (stitched paraphrases), inadequate paraphrase, missing citations, **self-plagiarism**. **Paraphrase** = fully restate + cite; **quote** = exact words in quotes + cite + page. When in doubt, cite.

**Turnitin** similarity index is not a plagiarism verdict - properly quoted/cited text matches legitimately; HEC thresholds guide but judgment rules. Keep drafts and notes as **evidence of authorship**.

**Fabrication/falsification** of data is the cardinal sin - never.

- **Every non-original idea gets a citation** - no exceptions.
- **Paraphrase + cite**; quote sparingly with page numbers.
- **Reference managers** (Zotero) eliminate formatting errors.
- **Keep an authorship trail**: notes, drafts, data files.

#### Example

Patchwriting: changing every third word of a source paragraph without citation is still plagiarism - restate the idea completely or quote it.

#### VERY IMPORTANT — Exam Pointer

Very important: distinguish paraphrasing, quoting and patchwriting with the citation each needs.

### Section Summary

Credit every idea; paraphrase honestly, quote exactly, and let tools check - not judge.

## Lecture 14: Research Ethics

### Learning Objectives

- Apply core ethical principles
- Navigate IRB approval
- Handle ethical dilemmas

### Key Terms & Definitions

#### Informed consent

Voluntary agreement based on full disclosure of the study.

#### Anonymity vs Confidentiality

No identifiers collected vs identifiers protected.

#### IRB/Ethics committee

Institutional board approving human-subject research.

**Principles** (Belmont): respect for persons (autonomy, consent), beneficence (maximize benefit, minimize harm), justice (fair selection of subjects). Operationalized: **informed consent**, voluntary participation, right to withdraw, **anonymity/confidentiality**, debriefing, data security.

**IRB approval** is required before collecting human data in most universities: submit protocol, instruments, consent forms; expedited vs full review by risk level. **Deception** needs strong justification + debriefing. Vulnerable populations (minors) need extra safeguards and guardian consent.

**Computing-specific:** scraping/API terms of service, user data privacy (PDPA/GDPR), dual-use concerns (security research - **responsible disclosure**), authorship ethics (who qualifies: substantial contribution; gift/ghost authorship is misconduct). **Conflicts of interest** must be declared.

- **Belmont trio:** respect, beneficence, justice.
- **Consent is a process**, not a form - ongoing and revocable.
- **Anonymize early:** separate identifiers from data at collection.
- **Authorship = substantial contribution** (conception, analysis, writing).

#### VERY IMPORTANT — Exam Pointer

Exam pointer: a researcher wants to scrape student Facebook posts for a study - list the ethical issues.

### Section Summary

Ethics protects participants and the researcher's credibility; get approval before data, not after.

## Lecture 15: Publishing and Research Impact

### Learning Objectives

- Choose publication venues
- Navigate peer review
- Understand impact metrics

### Key Terms & Definitions

#### Peer review

Independent expert evaluation before publication.

#### Impact factor

Journal-level citation metric (mean citations per article).

#### h-index

Researcher metric: h papers with at least h citations each.

**Venues:** in CS, top **conferences** (NeurIPS, ICSE, SIGCOMM) outrank most journals; elsewhere journals dominate. Check **HEC recognized** lists for Pakistan. Avoid **predatory journals** (spam invitations, fake impact factors, no real review - check Beall's criteria/ThinkCheckSubmit).

**Peer review:** single/double-blind; respond with a **rebuttal** addressing every comment (agree + change, or politely disagree with evidence). Rejection is normal - revise and resubmit elsewhere. **Open access** (gold/green) widens readership; watch APC costs.

**Metrics:** journal impact factor, **h-index**, i10-index, citations - useful but gameable; DORA warns against metric misuse. Real impact = uptake by researchers/practitioners. Build a profile: Google Scholar, ORCID,

ResearchGate.

- **ThinkCheckSubmit** before submitting anywhere unfamiliar.
- **Rebuttals**: number every reviewer point; show exact changes.
- **Predatory red flags**: guaranteed acceptance, broad scope spam, fake metrics.
- **ORCID + Scholar profile** from day one.

**VERY IMPORTANT — Exam Pointer** 

Very important: list five red flags of predatory journals.

## Section Summary

Publish where the community reads, survive review with grace, and measure impact honestly.

## Glossary of Key Terms

**FINER** — Feasible, Interesting, Novel, Ethical, Relevant - criteria for research questions.

**Paradigm** — Worldview guiding research: positivism, interpretivism, pragmatism.

**IMRaD** — Introduction, Methods, Results, and Discussion paper structure.

**p-value** — Probability of extreme data under the null hypothesis.

**Saturation** — Point where new data yields no new themes (qualitative sample-size rule).

**Triangulation** — Converging multiple methods/sources on one finding.

**Belmont principles** — Respect for persons, beneficence, justice.

**h-index** — h papers cited at least h times each.