



Conducting Rigorous and Scientific Research

Chang Xu

Department of Computer Science and Technology

Nanjing University

Nov 10, 2023

Outline



- I. Writing and Presentation
- II. Research Formulation
- III. Experimentation (main focus)



I. Writing and Presentation

1. Typo and grammar mistake
2. Sentence and convention
3. Exercise and discussion

General



- Writing
 - Make my paper error-free (readable)
 - E.g., no typo or grammar mistake
- Presentation
 - Make my paper *logically* written (readers can follow)
- Writing and presentation are different!
 - Writing first (own effort), presentation next (training)

Common Writing Mistakes (1)



■ Fragments

- *If* I had saved some money. I could have bought that new stereo.
- *Dancing* the night away in my new gown.
- I wanted a number of items. *Especially* lettuce and mushrooms.
- I visited the Philadelphia Museum of Art. *And* was most impressed with it.

Common Writing Mistakes (2)



- Run-together and comma-spliced sentences
 - My back hurts it is weak.
 - (Revised) My back hurts *because* it is weak.

 - My back hurts, it is weak.
 - (Revised) My back hurts, *for* it is weak.

Common Writing Mistakes (3)



- Subject and verb agreement
 - *They has worked* all day.
 - (Revised) *They have worked* all day.

 - *Each* of the nurses *are* present.
 - (Revised) *Each* of the nurses *is* present.

 - *Somebody* in the room *have taken* my wallet.
 - (Revised) *Somebody* in this room *has* taken my wallet.

Common Writing Mistakes (4)



- Ambiguous references
 - When *people* discuss their *problems*, *they* should be objective.
 - (Revised) *People* should be objective when *they* discuss their problems.
 - (Revised) The *problems which* people discuss should be objective.

Common Writing Mistakes (5)



- Problems with our “relatives”
 - I lost my *job, which* really angered me.
 - (Revised) When I lost my job, I was angry.
 - (Revised) Losing my job made my angry.

 - I had to *leave, which* was disturbing.
 - (Revised) Because I had to leave, I was disturbed.
 - (Revised) Having to leave disturbed me.

Common Writing Mistakes (6)



■ Dangling modifiers

- *Walking down the street*, a trash can blocked Joe's path.
- (Revised) *Walking down the street*, Joe could see a trash can blocking his path.
- (Revised) *While Joe was walking down the street*, a trash can blocked this path.

Common Writing Mistakes (7)



- Misplaced modifiers
 - Men and women who smoke *often* die of lung cancer.
 - (Revised) Men and women who smoke *excessively often* die of lung cancer.
 - (Revised) Men and women who smoke *will often die* of lung cancer.
 - (Revised) *Quite often*, men and women who smoke die of lung cancer.

Common Writing Mistakes (8)



- Faulty parallel structures
 - Joe wanted a *salary* higher than his brother *Bill*.
 - (Revised) Joe wanted a *salary* higher than his brother Bill's *salary*.
 - (Revised) Joe wanted a *salary* higher than *that* of his brother Bill.

More...



- You are supposed to have known all of these!
 - No reason to make such mistakes

- Do not ask me about English grammars
 - Ask yourself
 - Pick up CET-4/CET-6 and TOEFL/GRE materials

Sentences



- First
 - Write correct sentences

- For good presentation
 - Use *simple* sentences
 - Abandon *useless* (irrelevant) sentences

Writing Conventions (1)



- Avoid ambiguous words
 - E.g., if => whether

- Avoid offensive words
 - E.g., obviously, very, ...

- Do not put “and”, “also”, “but”, or “because” at the beginning of a sentence

Writing Conventions (2)



- Use terms/words consistently
 - E.g., We conducted an *experiment* ... This *evaluation* ...
 - E.g., Section 1 introduces ... Section 2 gives ... *We explain* ... in Section 3
- can not => cannot
- “e.g.,” “i.e.,” “, respectively”

Writing Conventions (3)



- A, B and C \Rightarrow A, B, and C
 - The group consists of two engineers, Jack and John.
 - The group consists of two engineers, Jack, and John.

Writing Conventions (4)



- Avoid passive tone
- Avoid subjective tone
 - E.g., I feel ..., I am confident ...
- Do not omit “that”
 - E.g., note *that* such a path is always executable

Exercises



- Describe an object in the classroom/meeting room
 - Requirement: error-free and following conventions
 - Word limit: 100

- Describe an activity you attended today/yesterday
 - Requirement: error-free and following conventions
 - Word limit: 100



II. Research Formulation

1. Definition before use
2. Research problem
3. Inadequacy of related work
4. Insight of my proposal
5. Exercise and discussion

General



- Formulating research and writing papers are always coupled together
- Implication
 - What I am going to write in papers should have been *already* considered when I conduct research
- Research formulation \leq paper presentation

Definition and Use



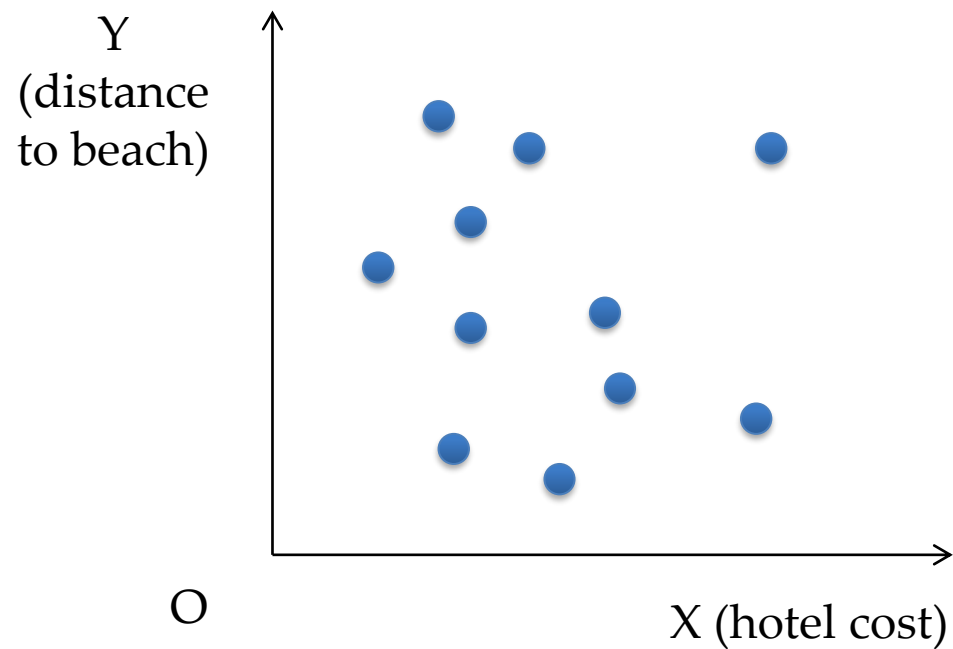
- Use consistent terms/words
 - E.g., experiment/evaluation, technique/approach, ...

- Concepts are defined before use
 - Each key concept must be defined, and non-key concepts should be discarded as many as possible
 - Each concept should carry the *same meaning* throughout the paper (so for variables)
 - Each defined concept should be *mathematically expressible or machine computable*

Example (1)



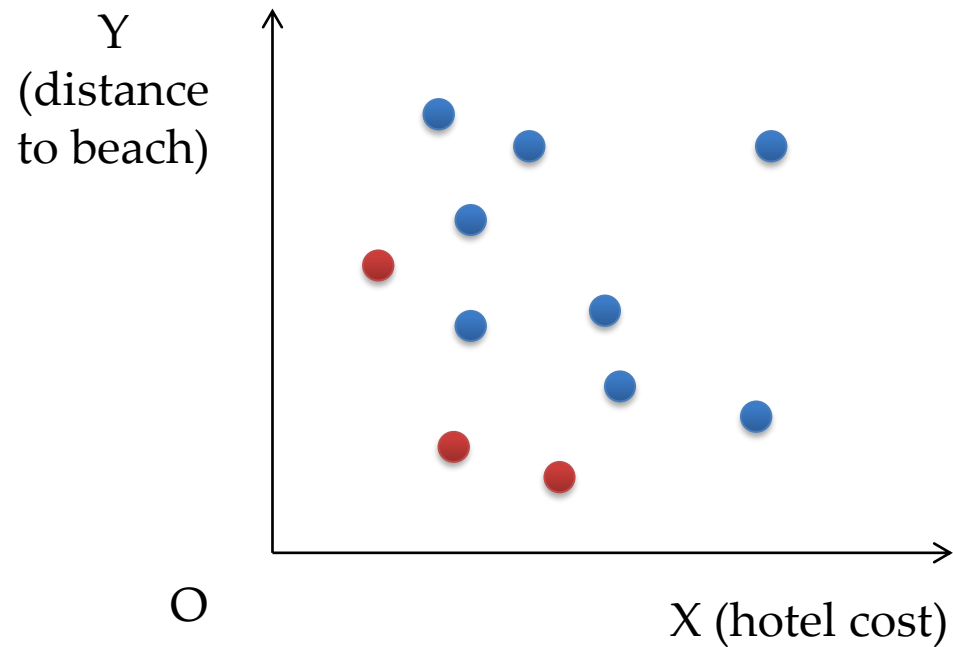
- Skyline points



Example (2)



- Skyline points (cont)



Exercise



- Consider one of your concepts and define it
 - Requirement: mathematically expressible or machine computable

Key Points



- I must have already considered
 - Research problem
 - Inadequacy of related work
 - Insight of my proposal

- Common mistake
 - They are what I have to consider when writing papers

Research Problem



- What I am going to address in this work

- Scope
 - What I can address (applicable scenarios)
 - What I cannot address (limitations)

- Never overclaim or claim imprecisely

Exercise



- Present your research problem

- Answer questions about its scope
 - What I can address
 - What I cannot address

Motivating Example



- Motivating example always with research problem
- Purpose
 - Give *concrete* evidence why we need new efforts
 - Not concrete => nothing (should be discarded)
- To where
 - Placed in the introduction or a separate section if it is large (when comparing several techniques)

Related Work



- Inadequacy of related work
 - Must be analyzed before proceeding to my proposal

- Requirements
 - The inadequacy should be demonstrated *concretely* using the motivating example or other strong evidence
 - The inadequacy should be discussed *precisely*

- Do not criticize others' work (say difference)

- Insight of my proposal
 - Key observations or efforts that make my proposal *able to (effectively) address* the research problem, while existing work cannot
 - Link it to a fundamental research problem

- This should not be implementation efforts
 - I am supposed to have made progress in research *methodology*

Common Mistakes



- Pick up a topic that seemingly has not been addressed by existing work
 - Why has it not been addressed before?
- Claim my work without justification
 - I need evidence to back me up
- Mixing framework and implementation
 - Must be *sound* at the framework (theory) level

Now



- You are supposed to
 - Have understood what a good presentation is
 - Work out how to get closer to this goal (long practice)

- Everybody has his (her) own way to go
 - We point out loopholes in your presentation
 - You fix them! And more iterations come ...

- Earn experience for yourself

Exercise



- Describe your work
 - Word limit: 150

- Requirement: good presentation
 - Self-explained
 - Research problem + inadequacy of existing work + insight of your proposal + evaluation results or plan



III. Experimentation

1. Questions and subjects
2. Experimental design
3. Threats to validity
4. Exercise and discussion

- Experimentation
 - Is not merely a description of the experimental procedure and a list of experimental results
 - Should have a careful *design* (questions and variables) and discuss potential *threats* (construct validity, internal validity, external validity, and conclusion validity)

Experimentation and Case Study



- Experimentation
 - In a lab environment
 - Variables (factors) can be isolated and *controlled*

- Case study
 - Under an industrial (real-world) setting
 - Hard to repeat

- We mostly conduct *controlled experiments*

Key Points (1)



- *Questions* to answer
 - Is Tool A *better* than Tool B?

Why would we expect it to be better?

Why do we need to know?

What will we do with the answer?

Better at *doing what?*

Better in *what way?*

Better in *what situations?*

E.g., study or physical exercise?

E.g., efficiency or durability?

E.g., in a classroom or dorm?

Key Points (2)



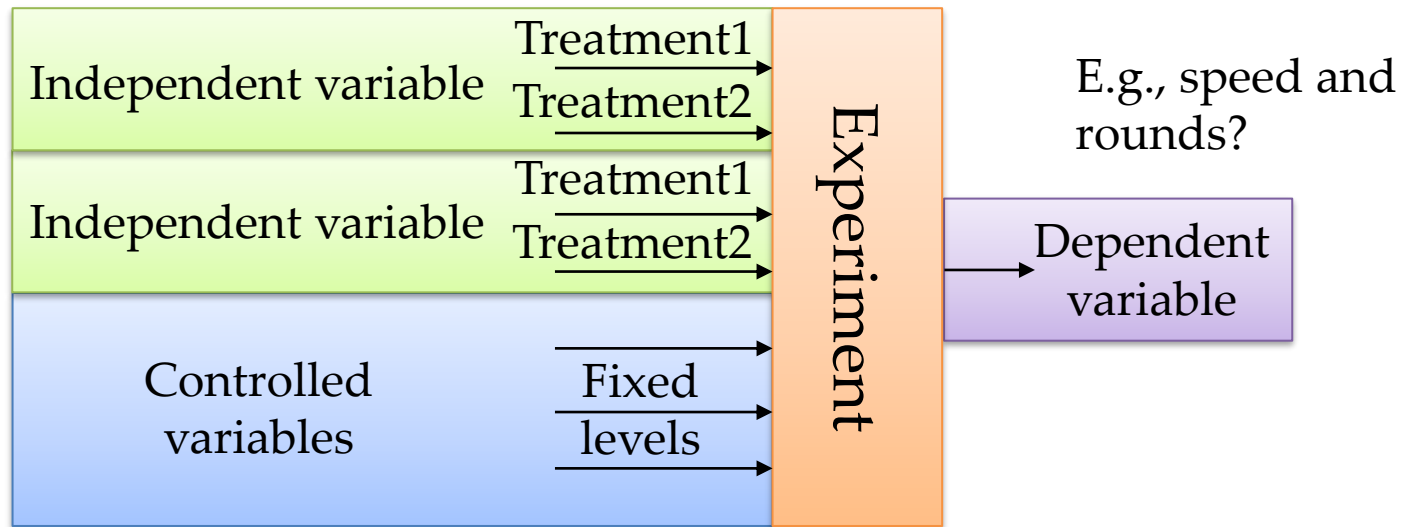
- *Subjects* selected
 - Sample of what population?
 - Consider the *representativeness*

- *Variables* and *threats* to validity
 - Variables: See the next page
 - Threats to validity: See an example

Variables

- Independent variables (factors)
- Dependent variables
- Controlled variables

E.g., mood?



E.g., speed and rounds?

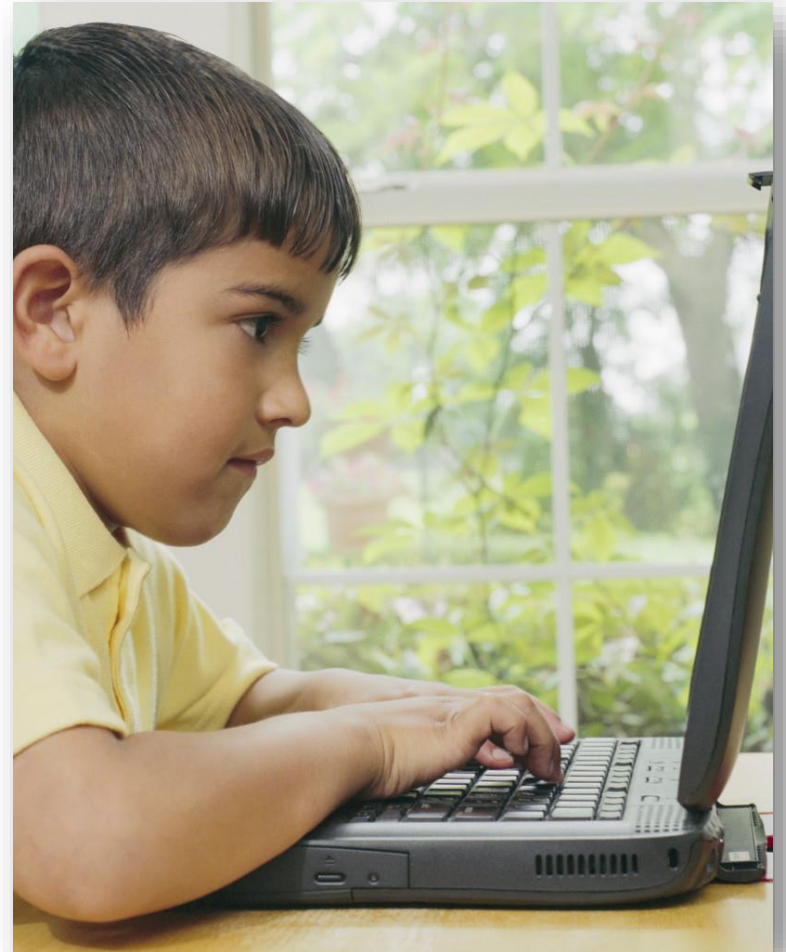
Dependent variable

E.g., location?

Example



- Name
 - Stuart Bean ("stu")
- Topic
 - Merging stakeholder views in model-driven development
- Status
 - 2 years into his PhD study
 - Has built a tool
 - Needs evaluation



Stu's Evaluation Plan



■ Experiments

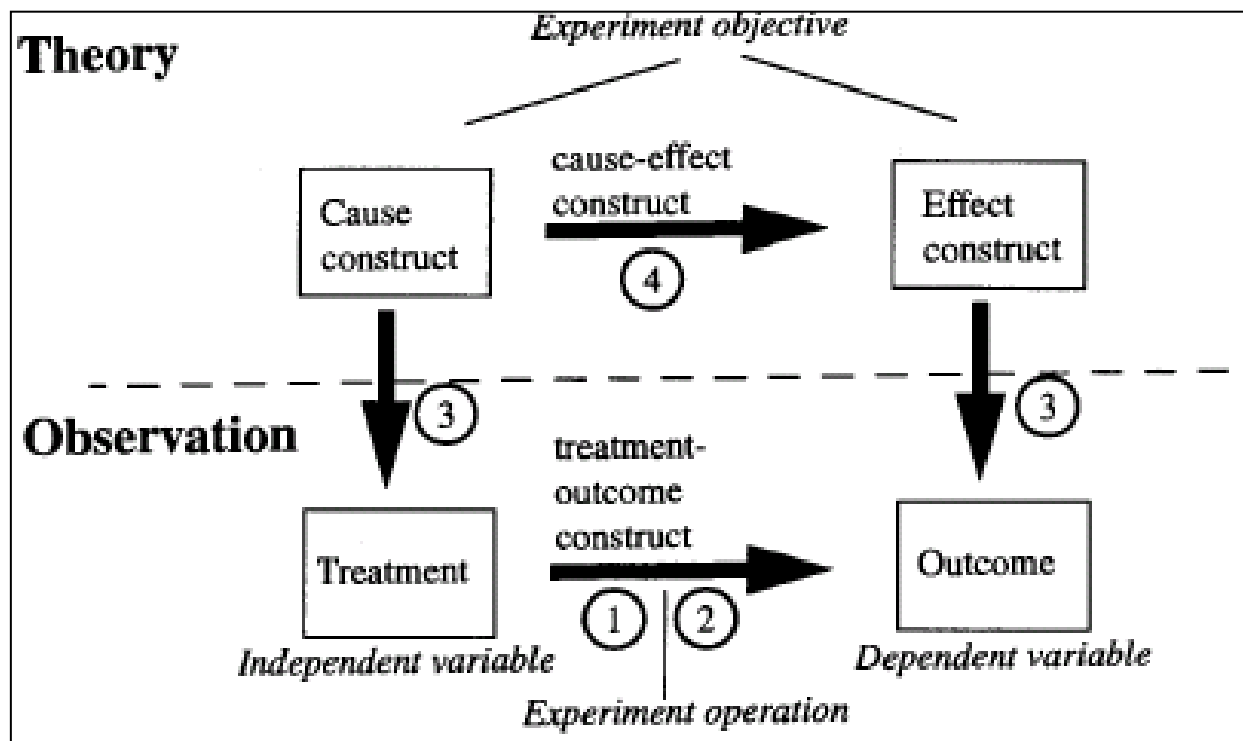
- Independent variable: Stu-merge vs. Rational Architect (RA)
- Dependent variables: correctness, speed, assessment
- Controlled variables: task (merging class diagrams from two different stakeholders' models), subjects (graduate students in software engineering)

■ Hypotheses

- H1: Stu-merge produces *correct* merges *more often* than RA
 - H2: Subjects produce merges *faster* with Stu-merge than RA
 - H3: Subjects *prefer* using Stu-merge to RA
- H1 accepted (strong evidence), H2 & H3 rejected

Threats to Validity Analysis

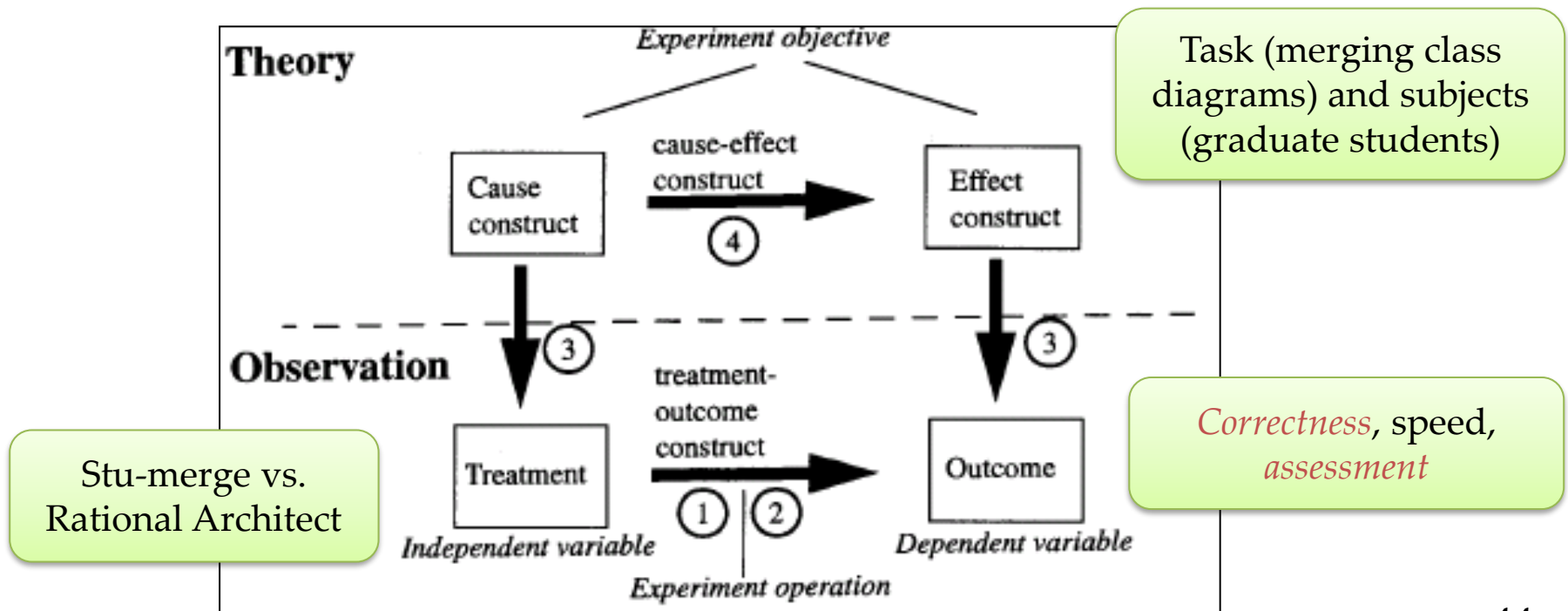
- ③ construct validity ② internal validity
- ④ external validity ① conclusion validity



Threats to Validity (1)

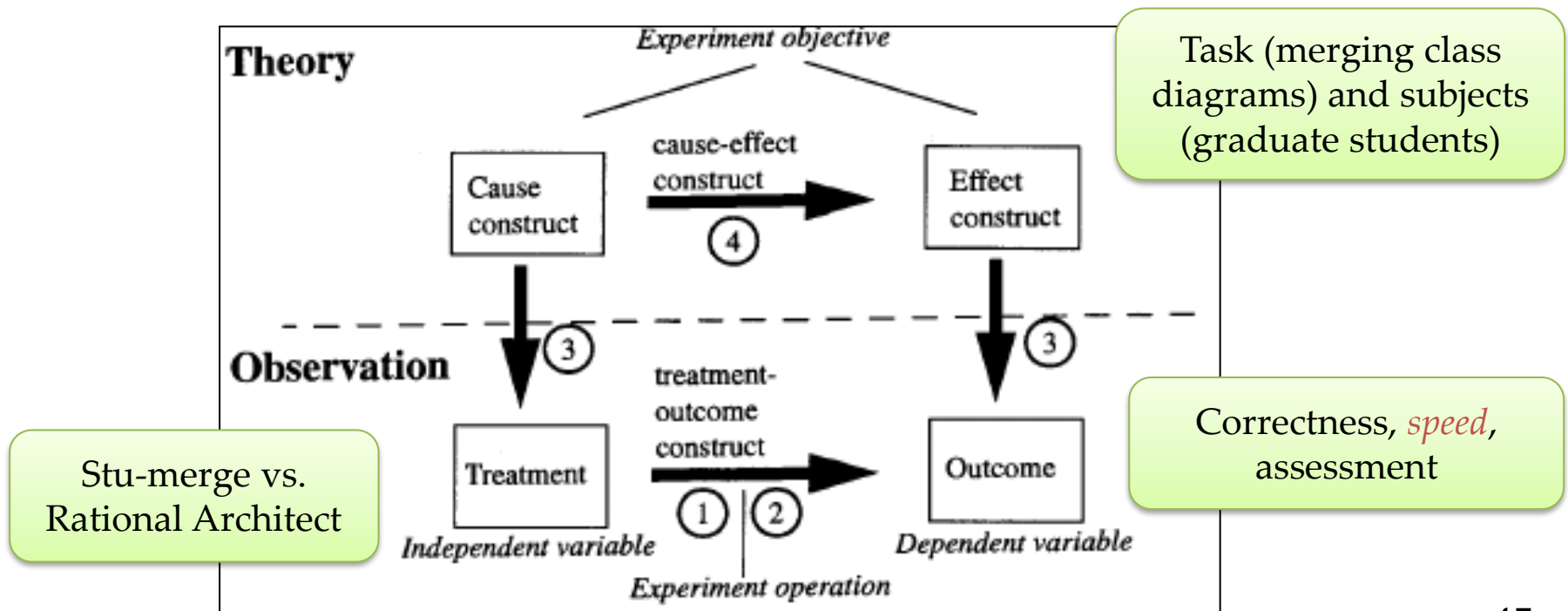


- ③ construct validity
 - What do we mean by a merge? What is correctness?
 - 1-5 point scale for subjective assessment - insufficient discriminatory power (both tools scored very low)



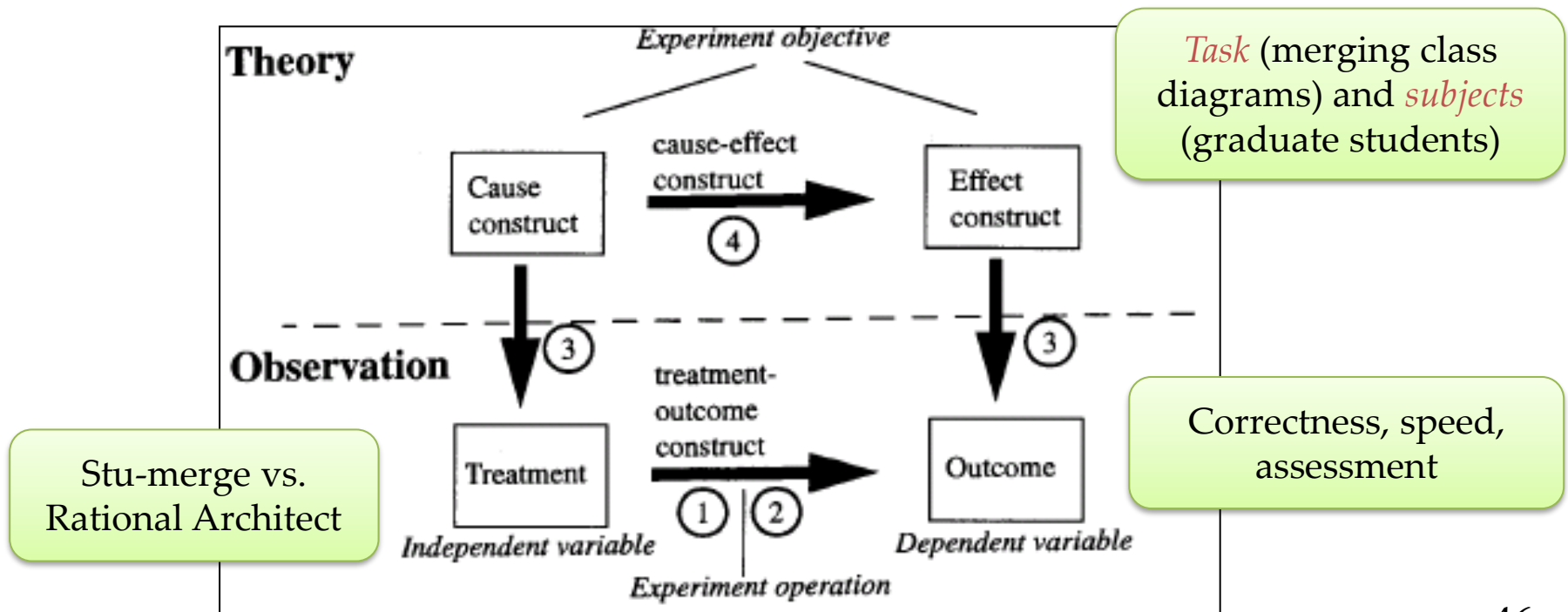
Threats to Validity (2)

- ② internal validity
 - *Confounding* variable: time taken to learn the tool (subjects were all familiar with RA, not with Stu-merge)



Threats to Validity (3)

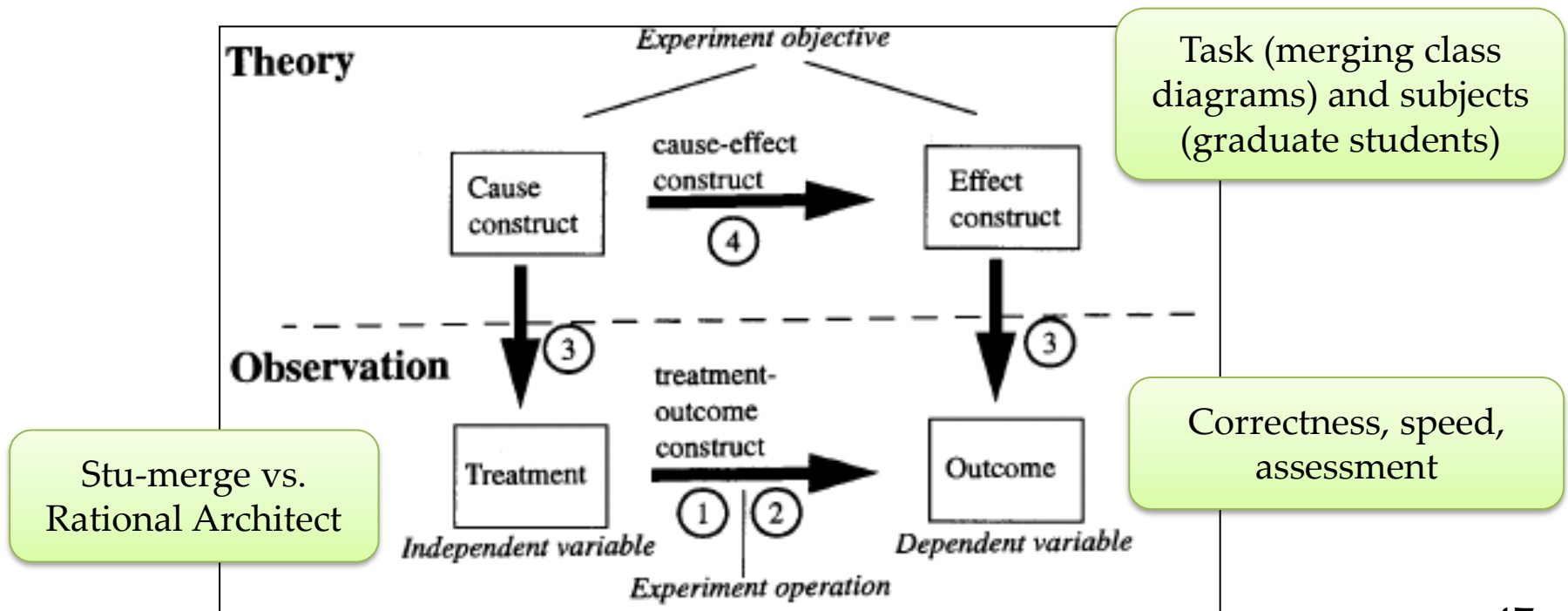
- ④ external validity (representativeness)
 - Task: class diagram models were of a toy problem
 - Subject: graduate students as sample of what population?



Threats to Validity (4)



- ① conclusion validity (theoretical reliability)
 - Bias: subjects knew Stu-merge was Stu's own tool



Exercise



- Describe your experimental design
 - *Questions* to answer
 - *Subjects* to select
 - *Independent* variables, *dependent* variables, and *controlled* variables (no *confounding* variable)

- Answer questions about
 - *Threats* to construct validity, internal validity, external validity, and conclusion validity
 - Why do they *not affect* your conclusion?

Referred Materials



- S.C. Cheung, “Empirical Experimentation”, COMP610F course slides, 2010.
- Steve Easterbrook, “Case Studies for Software Engineers”, FSE-DS keynote speech, 2006.
- Frances Kurilich and Helen Whitaker, “Re: Writing Strategies for Student Writers”, Harcourt College Pub, 1988.
- Xinming Wang, “Empirical Studies in Software Engineering”, group meeting talk, 2007.
- Tao Xie, “Common Technical Writing Issues” and “How to Write Research Papers”, HKUST-visit talk, 2007.