Lecture 9
VALIDITY THREATS

AISHA UMAIR

AIU@MMMI.SDU.DK

COURSE LEARNING OBJECTIVES

- The student is able to:
  - Independently find literature on a specific research topic and conduct systematic literature review
  - Identify qualitative and quantitative research approach and choose between them appropriately
  - Explain qualitative data collection and data analysis approaches
  - Explain quantitative data collection and data analysis approaches
  - Demonstrate ethical approach and research integrity in their own research
  - Identify common threats to quality of research and describe the validity and reliability of their own research
  - Present, both orally and in writing, the research process for a specific research topic, findings, and conclusion in a coherent and concise manner.
  - Write a scientific report covering a specific research topic using academic writing format and required content.

AGENDA

- What is a validity threat?
- Different threats to validity
- What to do about them?
- Exercise: (No submission required)

What is a Threat????

What is a Validity Threat????

WHAT IS A VALIDITY THREAT?

- Anything that can compromise the outcome of your study, e.g.,:
  - Inappropriate selection of statistical test
  - Incorrect measurement instrument in an experiment/survey
  - Incorrect formulation of interview/ survey questions
  - Biased interpretation of interview transcript
  - And many more .. if anything can go wrong it will

WHAT IS A VALIDITY THREAT?

- Incorrect formulation of interview/ survey questions

Examples:

Leading question: What problems do you have with the design team?
Fixed: How likely are you to recommend working with the design team?

Double-negative: Were the instructions not unclear?

Fixed: How would you rate the clarity of the instructions?

WHAT IS A VALIDITY THREAT?

- Biased interpretation of interview transcript

Some factors may bias the analysis of the transcripted interview (Tilley, 2003), for example

- Transcriber's awareness of the subject of the research,
- Predisposed attitudes towards the subject
- Difference in class, culture and language between the transcriber and the interviewees (Maclean, Meyer, & Estable, 2004).

EXAMPLE: *Interview of HIV Positive Men*

The research team read a transcript where the participant's statement was continually interrupted by his sniffling, indicated in the transcript by ((sniff)).

When the team met to discuss this transcript, the sniffling became confusing and the subject of some debate.

Some thought the participant was crying during the interview, whereas others made assumptions about drug use.

The confusion was settled when the interviewer explained that the participant was sick and his nose was running. (Oliver and colleagues (2005)

Source: https://www.scribbr.com/methodology/types-of-validity/#construct-validity

WHY DO WE DISCUSS VALIDITY THREATS?

WHY TO DISCUSS VALIDITY THREATS?

- Tell your audience, and yourself, that you understand that there are limitations to yourstudy.
- Tell your audience that your study is believable and trustworthy.
  - Research honesty and integrity
- Correctly argue the findings of your results without misleading the readers.
  - E.g., If your survey has a small number of respondents, do not claim that your study represents the general population

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 1 (Internal Validity)
- Category 2 (External Validity
- Category 3 (Construct Validity)
- Category 4 (Conclusion Validity)
- Category 5 (Descriptive Validity)
- Category 6 (Interpretive Validity)

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 1 (Internal Validity) - “Factor/Factors that might affect cause and effect relationships but is/are unknown to the researcher.”

Example:

You work for an e-commerce company and the company wants to increase revenue by improving the user experience of their customers when shopping online. You want to find out if item “quick view” feature would improve user experience and help to increase number of purchased items.

E.g., the impact of “quick view” and “no quick view”, but you did not screen your participants for visual impairment.

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

EXAMPLE - EXPERIMENT

Hypothesis: ***Drinking a cup of coffee improves memory.***

Sample: College-aged participants for morning and evening sessions at the laboratory
Treatment Group: Morning session participants
Control Group: Evening session participants

Once they arrive at the laboratory, the treatment group participants are given a cup of coffee to drink, while control group participants are given water. After coffee/water, you also give both groups memory tests. After analyzing the results, you find that the treatment group performed better than the control group on the memory test.

Source: https://www.scribbr.com/methodology/internal-validity/

How to check whether your study has internal validity?

- Your treatment and response variables change together.
- Your treatment precedes changes in your response variables
- No confounding or extraneous factors can explain the results of your study.

A confounding variable is a third variable that influences the relationship between independent and dependent variables by suggesting the presence of a spurious correlation.

Can you conclude that drinking a cup of coffee improves memory performance?

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 2 (External Validity) - “concerned with the ability to generalize the results”

- E.g., you conducted a case study on the implementation of DevOps in one robotics company in Odense. To what degree your findings would relevant to other robotics companies in Odense, or in Denmark. Or general implementation of DevOps in Denmark or even the world.

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

EXAMPLE

Hypothesis: ***People tend to perceive themselves as smarter than others in terms of academic abilities.***

Target population= 10,000 undergraduate students

Sample= 200 participants.

- *science and engineering majors;*
- *most of them are American,*
- *male,*
- *18–20 years old and*
- *from a high socioeconomic background*.

In a laboratory setting, you administer a mathematics and science test and then ask them to rate how well they think performed. You find that the average participant believes they are smarter than 66% of their peers.

Can you conclude that most people believe themselves to be much better than others at maths and science?

Source: https://www.scribbr.com/methodology/external-validity/

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 3 (Construct Validity) - “Concerned with whether we measured (e.g. quantitative measurement instrument) or captured (e.g. qualitatively in an interview), what we intend to in relation to our hypothesis or theory to test.”

E.g., is quick view feature representative of users’ convenience.

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

EXAMPLE 1:

If you develop a questionnaire to diagnose depression, you need to know:

- *Does the questionnaire really measure the construct of depression?*

Or

- *Is it actually measuring the respondent's mood, self-esteem, or some other construct?*

To achieve construct validity, you have to ensure that your indicators and measurements are carefully developed based on relevant existing knowledge.

The questionnaire must include only relevant questions that measure known indicators of depression.

Source: https://www.scribbr.com/methodology/types-of-validity/#construct-validity

EXAMPLE 2:

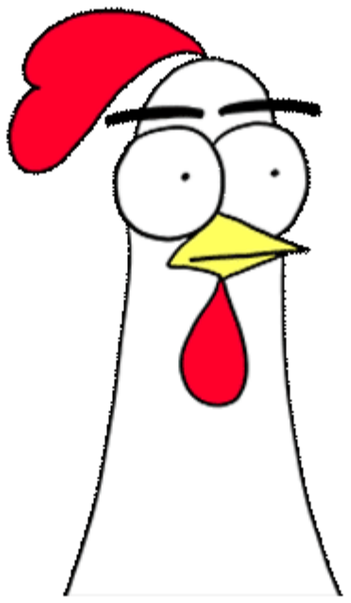
Consider the development of a software tool designed to assess the performance and security of web applications.

To ensure construct validity, it's crucial to ensure that the software tool accurately assesses the intended constructs of both performance and security. This means validating that the tool measures metrics such as response time, and vulnerabilities effectively.

EXAMPLE 3

Assume a researcher would like to assess the subject's experience in a programming language, measuring the number of courses at the university in computer science.

Do you think, the number of courses is the right measure of subject's experience?



EXAMPLE 3

- Assume a researcher would like to assess the subject's experience in a programming language, measuring the number of courses at the university in computer science.

- This may be a poor measure of the subject's experience in a programming language, , i.e. has poor construct validity.

- The number of years of practical use may be a better measure, i.e. has better construct validity

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 4 – (Conclusion Validity): “The degree to which conclusions/inferences we draw (e.g.about relationships between variables, or based on qualitative data) are reasonable.”
  - Establishing a correlation between two variables based on a very small sample.
  - Even in qualitative study, could two interviews really allow you to draw reasonable conclusion?

    OR

  - Is it reasonable to conclude based on the no. of observations that lack of motivation is related to constant pressure

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 5 – (Descriptive Validity): “Related to factual accuracy of the account/data; that is, the researchers are not making up or distorting the things they observed and it is expected to produce descriptively same accounts/data for the same event or situation”

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

EXAMPLE

The same list of primary studies will appear

Reliability concerns the degree of certainty with which a replication of this study, e.g., by a different set of researchers, would yield the same study outcome. As the search strategy, as well as the inclusion and exclusion criteria, is explicitly given in this study, the main reliability threat concerns the analysis resulting in the aggregation from reported effects of TDD to limiting factors. Particularly effects with low construct validity may be interpreted differently in replicated reviews. Hence, we have sought to describe the research process, including the data analysis, in a transparent manner.

Causevic, Adnan, Daniel Sundmark, and Sasikumar Punnekkat. "Factors limiting industrial adoption of test driven development: A systematic review." *2011 Fourth IEEE International Conference on Software Testing, Verification and Validation*. IEEE, 2011.

CATEGORIES OF VALIDITY THREATS (PETERSEN AND GENCEL, 2013)

- Category 6 - (Interpretive Validity): "Concerns with whether the inferences/ conclusions follow from the account (data), not biased by the researchers during analysis."
- More relevant in qualitative studies where you need to interpret the data.

K. Petersen and C. Gencel, "Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research," 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference o Software Process and Product Measurement, Ankara, 2013, pp. 81-89.

Break (10 min)

WHAT TO DO WITH VALIDITY THREATS?

- Prevent
- Mitigate
- Admit

PREVENT

- Ensure your instruments are well developed and verified for correctness.
  - Proof read instructions and questions
  - Ensure tools are working accordingly (audio recorders, or experiment objects)
- Pilot the study whenever possible
- Ensure that the data you collect is aligned with theory, if no theory, then it should answer the research question.
- Choose the analysis method that would suit the kind of data you would be collecting

PREVENT: Example

Validity Threats: *An Experimental Evaluation of Test Driven Development vs. Test-Last Development with Industry Professionals (Experiment)*

"In order to eradicate the selection threat to external validity, only professional Java software developers were included in the experiment.

Control group bias could be another threat to internal validity of the experiment that deals with the bias when the subjects belongs to the group that feels less important and they would have acted differently. This threat has been dismissed by creating two different experiment guidelines with different pre- and post-experiment questionnaires"

Munir, Hussan, et al. "An experimental evaluation of test driven development vs. test-last development with industry professionals." proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014.

MITIGATE

- Redo the study
- In case of little data points, carefully select the statistical methods. Avoid inferential statistics.
- Discard invalid answers or faulty data.
- Triangulate

TRIANGULATION!

- Taking different angles of a phenomenon, to widen the view and increase precision.
- Four types of triangulation (Stake, 1995):
  - **Data (source) triangulation**—using more than one data source:
    - multiple participants, multiple data source (interviews + document)
  - **Observer triangulation** —using more than one observer in the study.
  - **Methodological triangulation** —combining different types of data collection methods, e.g. qualitative and quantitative methods.
  - **Theory triangulation**— using alternative theories or viewpoints.
    - Use psychology to better understand code comprehension.

ADMIT/MITIGATE: Example

Validity Threats: *An Experimental Evaluation of Test Driven Development vs. Test-Last Development with Industry Professionals (Experiment)*

"Another threat relates to the selection of right assignment for the experiment. The bowling game used in the experiment may not be the most representative example of an industry project. However, due to the limited availability and time constraints it was not possible to conduct an experiment with an enterprise level application that requires significantly more effort from our subjects. Moreover, the bowling game has been used in the previous experiments related to TDD which makes our results comparable with them"

Munir, Hussan, et al. "An experimental evaluation of test driven development vs. test-last development with industry professionals." proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014.

ADMIT

- There are threats that cannot be prevented or mitigated. In that case you need to acknowledge that in the report.
  - In a retrospective study, participants might forget details.
  - In an experiment, if you have a homogeneous subjects (all female between 19-25), it may not be applicable to the wider population.

ADMIT: Example

Validity Threats: *Understanding the Efficacy of Test Driven Development (Interview)*

"One of the limitations relate to the fact that the data was obtained from semi-structured interviews only, with no triangulation from other data sources. Experts (e.g. Tellis, 1997; Yin, 2003) recommend that "data triangulation" in case study research is important to improve the validity of the research. This usually consists of a comparison between data obtained from different sources such as interviews, document analysis and observation"

"Data sources for this research are participants' perceptions collected through interviews, and it is possible for them to be interpreted differently to their original meanings by the researcher, thus threatening validity of the data. According to Yin (2003), relevant participants should review the data before they are recorded in the final report. However, due to time restrictions this activity was not conducted"

Li, Ling Angela. *Understanding the efficacy of test driven development*. Diss. Auckland University of Technology, 2009.

ADMIT: Example

Validity Threats: *A Comparative Case Study on the Impact of Test-Driven Development on Program Design and Test Coverage (Literature Review)*

"The use of students as study subjects places a question about the generalizability of the results.

The variety of the programming experiences in the case projects pose a threat to the internal validity of this study. In industry, teams usually have a mixed set of experience and skills, as many students work in industrial companies before their graduation. When doing industrial case studies, all these developers are usually regarded as professionals"

Siniaalto, Maria, and Pekka Abrahamsson. "A comparative case study on the impact of test-driven development on program design and test coverage." First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007). IEEE, 2007.

Mini Quiz 

An experiment high in internal validity identifies that the _____ caused changes in the _____

- ❑ Independent variable; dependent variable
- ❑ Dependent variable; independent variable
- ❑ All are correct
- ❑ None are correct

What kind of validity is threatened when we cannot generalize results of a study?

- ❑ external validity
- ❑ internal validity
- ❑ construct validity
- ❑ statistical validity

Internal validity is concerned with _____

- ❑ the generalizability of findings
- ❑ the validity of measures we used
- ❑ whether we can attribute effects to the independent variable
- ❑ the use of appropriate statistical analysis

IN-CLASS EXERCISE (NO SUBMISSION)

- Find other papers that are similar to the kind of project you are working on.
- Identify their validity threats!
- Reflect on your project, see if there are threats that could be relevant to you.