

Internal validity

- Defending against sources of bias which would affect the cause-effect process being studied by introducing covert variables.
- When there is lack of internal validity, variables other than the IV being studied may be responsible for part or all of the observed effect on the dependent variable(s).
- An experiment is internally valid to the extent that it shows a cause-effect relationship between the independent and dependent variables.

Sources of threat to internal validity

- Selection biases
- History
- Maturation

Sources of threat to internal validity, cont.

- Repeated Testing
- Instrumentation

Sources of threat to internal validity, cont.

- Regression to the Mean
 - Subjects with extreme scores on a first measure of the DV tend to have scores closer to the mean on a second measure.
 - Example:
 - A person who scored 750 out of a possible 800 on the quantitative portion of the GRE takes the GRE again.
 - Assuming the second test is the same difficulty as the first and there is no learning or practice effect, what score would you expect the person to get on the second test?

Regression to the mean

- The person is more likely to score below 750 than above 750.
- Why? Some portion of scores on the test are likely due to chance factors.
- If you picked the person on the basis of extreme scores, it is more likely that the extreme nature of the score is partly the result of chance.

Regression to the mean, cont.

- Consider the ways someone could score 750.
 - (1) "true" score is 750 & they had exactly average luck
 - (2) "true" score is < 750 & they had better than average luck
 - (3) "true" score is > 750 & they had worse than average luck
- There are very few people with "true" scores above 750 (roughly 6 in 1,000); there are many more people with true scores between 700 and 750 (roughly 17 in 1,000).
- Thus it is more likely that someone scoring 750 is from the former group and was lucky than from the latter group and was unlucky.

Regression to the mean in the clinic

- Which kids get referred to a clinic?
 - Those who do extremely poorly.
 - Some may have done poorly from chance factors – and would improve if retested EVEN WITHOUT CLINICAL INTERVENTION.
- So if you do clinical intervention and then retest, you do not know that the improvement is because of the intervention – it might be because of regression to the mean alone.

Another example....

- The *Sports Illustrated* jinx:
 - After appearing on the cover of the magazine, players' averages tend to drop...

Sources of threat to internal validity, cont.

- Experimental Mortality
- Selection-Maturation Interaction
- Experimenter Bias

Some examples.... background

- The goal in this study was to teach children, upon being enticed by stranger, to say "No, I have to go ask my teacher" and then to quickly run away. The experimenter wanted to evaluate the effectiveness of a 20-minute interactive videotape program.
- In the Experimental condition, children viewed the interactive video; in the Control condition, children viewed a cartoon of equal duration having nothing to do with child abduction.
- A Generalization Probe tested what each child would do in a potential abduction situation.
- For purposes of data analysis:
 - A score of zero was given if a child went with the abductor
 - A score of one if the child stayed nearby with no refusal
 - A score of two if the child stayed nearby but verbally refused
 - A score of three if the child ran away with no refusal
 - A score of four if the child ran away and verbally refused.

All versions of this study from psych.athabasca.ca/html/Validity/

Case 1

- It so happened that there were an equal number of boys and girls in the classes, so for convenience the boys were assigned to the Control Group and the girls to the Experimental Group.
- One day at school, the boys were told to go to one room and the girls to another room, where they were exposed to their respective conditions. Two days later, the Generalization Probe was conducted.
- The mean score for children in the Control Group was 1.2 and the mean score for children in the Experimental Group was 3.4.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 2

- Due to time constraints, the experiment was run over four consecutive days.
 - On Day 1, children viewed the 20-minute cartoon (Control condition).
 - On Day 2, the Generalization Probe was conducted.
 - On Day 3, the children were exposed to the 20-minute interactive video (Experimental condition).
 - Finally, on Day 4, a second Generalization Probe was conducted.
- The mean score for children on the first Generalization Probe was 1.2 and their mean score on the second Generalization Probe was 3.4. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 3

- The name of each child in the classes was written on a separate slip of paper; all the slips were put in a bowl and mixed up thoroughly. Students were assigned to the Experimental Group and to the Control Group alternately as their names were pulled out of the bowl one at a time.
- One day at school, the children in the Control Group were told to go to one room and the children in the Experimental Group to another room, where they were exposed to their respective conditions.
- Immediately afterwards, while walking back to their regular classroom, all the children in the Control Group saw a man laughing and joking with their school principal.
- Two days later, the Generalization Probe was conducted, during which many of the Control Group children recognized the stranger as the man who made their principal laugh. The mean score for children in the Control Group was 1.2 and the mean score for children in the Experimental Group was 3.4. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 4

- The name of each child in the classes was written on a separate slip of paper. All the slips were put in a bowl and mixed up thoroughly. Students were assigned to the Experimental Group and to the Control Group alternately as their names were pulled out of the bowl one at a time.
- One day at school, the children in the Control Group were told to go to one room and children in the Experimental Group to another room, where they were exposed to their respective conditions. Two days later, the Generalization Probe was conducted.
- For ease of record keeping, all Control Group children were tested first, then all the Experimental Group children. The student teacher scored children's responses to the confederate's lures. In the beginning, he hid indoors and strained to see and hear through an open window; later on, he discovered he could see and hear better by hiding outside and peeking around a corner.
- The mean score for children in the Control Group was 1.2 and the mean score for children in the Experimental Group was 3.4. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 5

- During a class early in the school year, the children viewed the 20-minute cartoon (Control condition). Two days later, the Generalization Probe was conducted.
- The experimenter fell ill soon afterwards, and so it wasn't until a class late in the school year that the children viewed the 20-minute interactive video (Experimental condition).
- Two days after that, a second Generalization Probe was conducted.
- The mean score for the children on the first Generalization Probe was 1.2 and their mean score on the second Generalization Probe was 3.4. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 6

- One day at school, the children viewed the 20-minute cartoon (Control condition).
- Two days later, the Generalization Probe was conducted. Then, in a class the following week, the children viewed the 20-minute interactive video (Experimental condition).
- The plan was to administer a second Generalization Probe two days after that. However, at this point, the experimenter realized that she had insufficient funding to complete the study and would only be able to retest ten children.
- She selected the ten poorest performing children on the first Generalization Probe, the mean score of which was 0.1. Their mean score on the second Generalization Probe was 2.5. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 7

- There were an equal number of boys and girls, so for convenience the boys were assigned to the Control Group and the girls to the Experimental Group.
- During a class early in the school year, a Generalization Probe was conducted for all children. The experimenter fell ill soon afterwards, and so it wasn't until a class late in the school year that the children in the comparison groups were separated, with the control children viewing the 20-minute cartoon and the experimental children viewing the 20-minute interactive video.
- Two days after that, a second Generalization Probe was conducted. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.

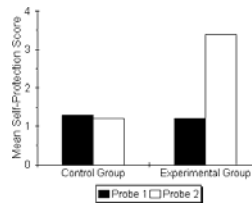


Figure Caption. The mean self-protection score on the two Generalization Probes for Control Group children and Experimental Group children.

All versions of this study from psych.athabasca.ca/html/Validity/

Case 8

- The name of each child in the classes was written on a separate slip of paper. All the slips were put in a bowl and mixed up thoroughly. Students were assigned to the Experimental Group and to the Control Group alternately as their names were pulled out of the bowl one at a time.
- One day at school, the children in the Control Group were told to go to one room and children in the Experimental Group to another room, where they were exposed to their respective conditions.
- Some of the children in the Experimental Group appeared bored by the interactive video, became disruptive, and were removed from the room. Two days later, the Generalization Probe was conducted.
- The mean score for children in the Control Group was 1.2 and the mean score for the remaining children in the Experimental Group was 3.4. We conclude that the 20-minute interactive video improved the children's self-protection skills in a potential abduction situation.
- What is the problem?

All versions of this study from psych.athabasca.ca/html/Validity/

Case 9

- The name of each child in the classes was written on a separate slip of paper. All the slips were put in a bowl and mixed up thoroughly. Students were assigned to the Experimental Group and to the Control Group alternately as their names were pulled out of the bowl one at a time.
- One day at school, children in the Control Group were told to go to one room and children in the Experimental Group to another room, where they were exposed to their respective conditions. Two days later, the Generalization Probe was conducted. For ease of record keeping, all the Control Group children were tested first, then all the Experimental Group children.
- Both the student teacher, who recorded how the children responded to the confederate's lures, and the confederate who presented the lures, were heavily involved in the production of the interactive video, and both of them strongly believed in its efficacy.
- The mean score for children in the Control Group was 1.2 and the mean score for children in the Experimental Group was 3.4. We can conclude that the 20-minute interactive video was effective in changing what the children did in a potential abduction situation.

All versions of this study from psych.athabascau.ca/html/Validity/

Is there internal validity?

- Does modeling improve an infant's ability to make a pincer grasp?
- 8-month-old infants were observed in 2 1-hr sessions
- Session 1: infants' parents presented them with a pea every 5 min. along with a verbal prompt to pick it up.
- Session 2 (2 months later): same thing, but also modeled the correct behavior immediately prior to the verbal prompt.
- Infants used the appropriate motor response 10% of time in session 1 and 45% of the time in the session 2.
- Thus, modeling increased the likelihood with which **infants made a neat pincer grasp.**

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #2

- Wilma's parents and teacher were concerned about her noncompliance, so they consulted a behavior therapist.
- They decided to first gauge the extent of the problem.
- For the next week, her teacher recorded all instances of the target behavior at school and her parents recorded all instances at home.
- Noncompliance was defined as a failure to respond to an instruction within 10 seconds.
- Wilma was four times more likely to obey the instruction at home than at school.

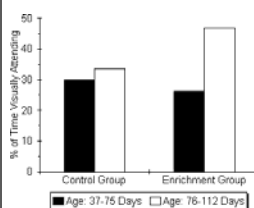
Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #3

- Does experience affect early human behavior?
- Infants in an institution were randomly assigned to a Control Group or an Enrichment Group (EG).
- Target behavior: visual attention
 - “The state in which the infant's eyes are more than half open, their direction of gaze shifting within 30 seconds.”
- Once a week, infants were observed for 3 hours in their cribs.
- Infants in EG were provided with extra mobility (3 additional periods per day of being placed in the prone position) & enhanced visual surroundings (a **colorful mobile suspended over the cribs**).

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #3, cont.



- We can conclude that the extra experiences of children in the Enrichment Group improved their visual attention.

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #4

- Psychoanalysts at two different hospitals were asked to judge the well-being of a young man being interviewed on videotape.
- By the flip of a coin, psychoanalysts at a publicly-funded hospital were assigned to the Normal Group (NG) and psychoanalysts at a privately-funded hospital were assigned to the Abnormal Group (AG).
 - Doctors in NG were told that the man was a job applicant.
 - Doctors in AG were told that he was a patient.
- Mean adjustment ratings: 7 out of 8 for NG, 3.5 out of 8 by AG.
- We can conclude that psychoanalysts' ratings were affected by the label used to describe the young man.

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #5

- Andy swore a lot, a bad habit he hoped to kick.
- His wife, Betty, recently graduated with a degree in behavior therapy.
- She suggested that Andy record every time he used profanity throughout the day by transferring small beads from one pocket to another, and then counting all the beads transferred at the end of the day.
- "Do this for the first week," she said, "and we'll see how often you typically swear." "Then, the next week continue counting, but we'll add a new twist: I will only cook dinner on days when you have sworn 50% less than your average daily count in the first week."

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #5, cont.

- Andy, who had never been supportive of his wife's choice of profession, was convinced that this was not going to work.
- In fact, he was extra vigilant during that second week just to make sure he didn't miss any opportunity to catch himself being profane.
- Sure enough, his daily rate of swearing remained constant throughout the two weeks despite missing dinner every night during that second week.

Example from psych.athabascau.ca/html/Validity/

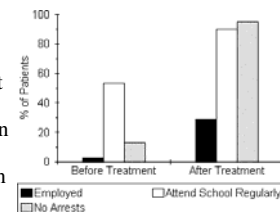
Is there internal validity? #6

- When juveniles in the care of foster parents get into extreme trouble with the law, they are sent to Correction House, a rehabilitative group home.
- Correction House has a reputation for treating "the worst of the worst," a last stop before imprisonment.
- The rehab program consists of a token economy
 - points are earned for appropriate social behaviors and lost for inappropriate behaviors
 - Points can be exchanged for backup reinforcers.
- Collected information on patients one month before entering the program and one month after leaving it.

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #6, cont.

- The typical patient stay at Correction House was about six months.
- We can conclude that the rehabilitative program at Correction House improved the juveniles' behavior on all three dependent measures.



Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #7

- Investigated a new treatment regimen for problem drinking, based on use of disulfiram.
 - Severe adverse physical reactions result from drinking alcohol after taking disulfiram.
- Outpatients of a rural community alcoholism treatment clinic were randomly assigned to either a Traditional Group (TG) or a Disulfiram Group (DG).
- TG patients attended sessions devoted to education about alcoholism.
- DG patients did likewise, but also took disulfiram.
 - Intervention stressed the importance of taking disulfiram at a set time & place and in the company of another

Example from psych.athabascau.ca/html/Validity/

Is there internal validity? #7, cont.

- 95% of TG patients completed the program
- 60% of DG patients completed the program
- Six months following treatment, patients in TG reported being abstinent on 15 of the previous 30 days while patients in DG reported being abstinent on 25 of these days.
- We can conclude that the Disulfiram treatment was more effective than the Traditional treatment.

Example from psych.athabascau.ca/html/Validity/

External validity

- Generalizeability.
- The degree to which the conclusions in your study would hold for other persons in other places and at other times.
- "To what population does the researcher wish to generalize his/her conclusions, and is there something unique about the study sample's subjects, the place where they lived/worked, the setting in which they were involved, or the times of the study, which would prevent valid generalization?"

Threats to external validity

- An explanation of how you might be wrong in making a generalization.
- There are three major threats to external validity because there are three ways you could be wrong -- people, places or times.

Problems with external validity

- If your participants are not similar to the population to which you want to generalize, meaningful generalization will be limited.
- Pretesting subjects may cause them to react more/less strongly to the treatment than they would have had they not experienced the pretest.
- Performance of subjects may be more a product or reaction to the experimental setting than it is to the independent variable.
- With multiple treatments/interventions, the early treatments may have a cumulative effect on the subjects' performance.