

Internal Validity

Martyn Shuttleworth 180.2K reads

Internal validity is a crucial measure in quantitative studies, where it ensures that a researcher's experiment design closely follows the principle of cause and effect.

Internal validity is an important consideration in most scientific disciplines, especially the social sciences.



What is Internal Validity?

The easy way to describe internal validity is the confidence that we can place in the cause and effect relationship in a study. The key question that you should ask in any experiment is:

“Could there be an alternative cause, or causes, that explain my observations and results?”

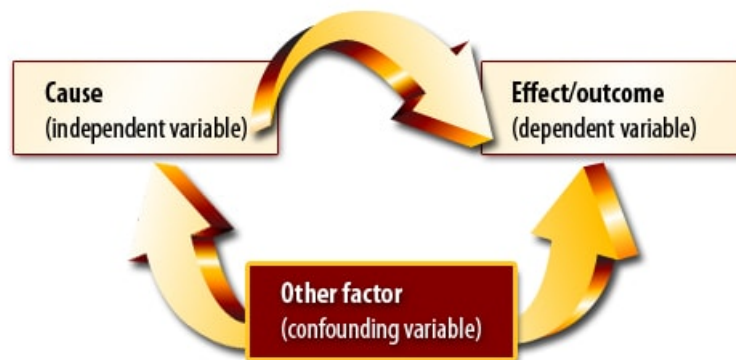
Looking at some extreme examples, a physics experiment into the effect of heat on the conductivity of a metal has a high internal validity.

The researcher can eliminate almost all of the potential confounding variables ^[1] and set up strong controls ^[2] to isolate other factors.

At the other end of the scale, a study into the correlation ^[3] between income level and the likelihood of smoking has a far lower internal validity.

A researcher may find that there is a link between low-income groups and smoking, but cannot be certain that one causes the other.

Social status, profession, ethnicity, education, parental smoking, and exposure to targeted advertising are all variables [4] that may have an effect. They are difficult to eliminate, and social research can be a statistical [5] minefield for the unwary.



Internal Validity vs Construct Validity

For physical scientists, construct validity [6] is rarely needed but, for social sciences and psychology, construct validity is the very foundation of research.

Even more important is understanding the difference between construct validity and internal validity, which can be a very fine distinction.

The subtle differences between the two are not always clear, but it is important to be able to distinguish between the two, especially if you wish to be involved in the social sciences, psychology and medicine.

Internal validity only shows that you have evidence to suggest that a program or study had some effect on the observations and results.

Construct validity determines whether the program measured the intended attribute.

Internal validity says nothing about whether the results were what you expected, or whether generalization [7] is possible.

For example, imagine that some researchers wanted to investigate the effects of a computer program against traditional classroom methods for teaching Greek.

The results showed that children using the computer program learned far more quickly, and improved their grades significantly [8].

However, further investigation showed that the results [9] were not due to the program itself, but due to the Hawthorne Effect [10]; the children using the computer program felt that they had been singled out for special attention. As a result, they tried a little harder, instead of staring out of the window.

This experiment ^[11] still showed high internal validity, because the research manipulation ^[12] had an effect.

However, the study had low construct validity, because the cause was not correctly labeled. The experiment ^[13] ultimately measured ^[14] the effects of increased attention, rather than the intended merits of the computer program.

How to Maintain High Confidence in Internal Validity?

It is impossible to maintain 100% confidence in any experimental design ^[15], and there is always the chance of error ^[16].

However, there are a number of tools that help a researcher to oversee internal validity and establish causality ^[17].

Temporal Precedence

Temporal precedence is the single most important tool for determining the strength of a cause and effect relationship. This is the process of establishing that the cause did indeed happen before the effect, providing a solution to the chicken and egg problem.

To establish internal validity through temporal precedence, a researcher must establish which variable came first.

One example could be an ecology study, establishing whether an increase in the population of lemmings in a fjord in Norway is followed by an increase in the number of predators.

Lemmings show a very predictable population cycle, which steadily rises and falls over 3 to 5 year cycle. Population estimates show that the number of lemmings rises due to an increase in the abundance of food.

This trend is followed, a couple of months later, by an increase in the number of predators, as more of their young survive. This seems to be a pretty clear example of temporal precedence; the availability of food for the lemmings dictates numbers. In turn, this dictates the population of predators.

STOP!

Not so fast!

In fact, the predator/prey relationship is much more complex than this. Ecosystems rarely contain simple linear relationships ^[18], and food availability is only one controlling factor.

Turning the whole thing around, an increase in the number of predators may also control the lemming population. The predators may be so successful that the lemming population plummets and the predators starve, through limiting their own food supply.

What if predators turn to an alternative food supply when the number of lemmings is low? Lemmings, like many rodents, show lower breeding success during times of high population.

This really is a tough call, and the only answer is to study previous research. Internal validity is possibly the single most important reason for conducting a strong and thorough literature review [19].

Even with this, it is often difficult to show that cause happens before effect, a fact that behavioral biologists and ecologists know only too well.

By contrast, the physics experiment is fairly easy - I heat the metal and conductivity increase/decreases, providing a simpler view of cause and effect and high internal validity.

An Example of a Lemming Study [20]

Covariation of the Cause and Effect

Covariation of the cause and effect is the process of establishing that there is a cause and effect to relationship between the variables. It establishes that the experiment or program had some measurable effect, whatever that may be.

For example, in the study of Greek learning, the results showed that the group with the computer package performed better than those without.

This can be summed up as:

If you use the program, there is an outcome.
Without the program, there is no outcome.

This does not need to be an either/or relationship and it could be:

More of the program equals more of the outcome.
Less of the program equals less of the outcome.

This seems pretty obvious, but you have to remember the basic rule of internal validity. Covariation of the cause and effect cannot explain what causes the effect, or establish whether it is due to the expected manipulated variable [12] or to a confounding variable [1].

It does, however, strengthen the internal validity of the study.

Establishing Causality Through a Process of Elimination

Establishing causality through elimination is the easiest way to prove that an experiment has high internal validity.

As with the lemming example, there could be many other plausible explanations for the apparent causal link between prey and predator.

Researchers often refer to any such confounding variable [1] as the 'Missing Variable,' an unknown factor that may underpin the apparent relationship.

The problem is, as the name suggests, that the variable is missing, and trying to find it is almost impossible. The only way to nullify it is through strong experimental design ^[21], eliminating confounding variables and ensuring that they cannot have any influence.

Randomization, control groups ^[22] and repeat experiments are the best way to eliminate these variables and maintain high validity.

In the lemming example, researchers use a whole series of experiments, measuring predation rates, alternative food sources and lemming breeding rates, attempting to establish a baseline.

Internal Validity - the Final Word

Just to leave you with an example of how difficult measuring internal validity can be:

In the experiment where researchers compared a computer program for teaching Greek against traditional methods, there are a number of threats to internal validity.

- The group with computers feels special, so they try harder, the Hawthorne Effect ^[10].
- The group without computers becomes jealous, and tries harder to prove that they should have been given the chance to use the shiny new technology.
- Alternatively, the group without computers is demoralized and their performance suffers.
- Parents of the children in the computerless group feel that their children are missing out, and complain that all children should be given the opportunity.
- The children talk outside school and compare notes, muddying the water.
- The teachers feel sorry for the children without the program and attempt to compensate, helping the children more than normal.

We are not trying to depress you with these complications, only illustrate how complex internal validity can be.

In fact, perfect internal validity is an unattainable ideal, but any research design ^[23] must strive towards that perfection.

For those of you wondering whether you picked the right course, don't worry. Designing experiments ^[15] with good internal validity is a matter of experience, and becomes much easier over time.

For the scientists who think that social sciences are soft - think again!

Source URL: <https://explorable.com/internal-validity>

Links

- [1] <https://explorable.com/confounding-variables>
- [2] <https://explorable.com/controlled-variables>
- [3] <https://explorable.com/correlation-and-causation>
- [4] <https://explorable.com/research-variables>
- [5] <https://explorable.com/statistics-tutorial>
- [6] <https://explorable.com/construct-validity>
- [7] <https://explorable.com/what-is-generalization>
- [8] <https://explorable.com/significance-test>
- [9] <https://explorable.com/statistically-significant-results>

- [10] <https://explorable.com/hawthorne-effect>
- [11] <https://explorable.com/experimental-research>
- [12] <https://explorable.com/independent-variable>
- [13] <https://explorable.com/conducting-an-experiment>
- [14] <https://explorable.com/scientific-measurements>
- [15] <https://explorable.com/design-of-experiment>
- [16] <https://explorable.com/systematic-error>
- [17] <https://explorable.com/causal-reasoning>
- [18] <https://explorable.com/linear-relationship>
- [19] <https://explorable.com/what-is-a-literature-review>
- [20] <http://pubs.aina.ucalgary.ca/arctic/Arctic50-4-377.pdf>
- [21] <https://explorable.com/true-experimental-design>
- [22] <https://explorable.com/scientific-control-group>
- [23] <https://explorable.com/research-designs>