

## Criterion Validity

Martyn Shuttleworth 110.6K reads

Criterion validity assesses whether a test reflects a certain set of abilities.

To measure the criterion validity of a test, researchers must calibrate it against a known standard or against itself.

Comparing the test with an established measure is known as concurrent validity; testing it over a period of time is known as predictive validity.

It is not necessary to use both of these methods, and one is regarded as sufficient if the experimental design is strong.

One of the simplest ways to assess criterion related validity is to compare it to a known standard.

A new intelligence test, for example, could be statistically analyzed against a standard IQ test; if there is a high correlation between the two data sets, then the criterion validity is high. This is a good example of concurrent validity, but this type of analysis can be much more subtle.



The banner features the Explorable logo at the top center, with the text "Quiz Time!" in a stylized font below it. Below the logo are three quiz cards, each with a different image and title. The first card shows a pair of red roller skates on a wooden deck, titled "Quiz: Psychology 101 Part 2". The second card shows a fan of colorful pencils, also titled "Quiz: Psychology 101 Part 2". The third card shows a Ferris wheel at sunset, titled "Quiz: Flags in Europe". To the right of the cards is a link that says "See all quizzes =>".

**EXPLORABLE**  
*Quiz Time!*

 Quiz: Psychology 101 Part 2

 Quiz: Psychology 101 Part 2

 Quiz: Flags in Europe

[See all quizzes =>](#)

## An Example of Criterion Validity in Action

A poll company devises a test that they believe locates people on the political scale, based upon a set of questions that establishes whether people are left wing or right wing.

With this test, they hope to predict how people are likely to vote. To assess the criterion validity <sup>[1]</sup> of the test, they do a pilot study <sup>[2]</sup>, selecting only members of left wing and right wing political parties.

If the test has high concurrent validity <sup>[3]</sup>, the members of the leftist party should receive scores that reflect their left leaning ideology. Likewise, members of the right wing party should receive scores indicating that they lie to the right.

If this does not happen, then the test is flawed and needs a redesign. If it does work, then the researchers can assume that their test has a firm basis, and the criterion related validity is high.

Most pollsters would not leave it there and in a few months, when the votes from the election were counted, they would ask the subjects how they actually voted.

This predictive validity <sup>[4]</sup> allows them to double check their test, with a high correlation again indicating that they have developed a solid test of political ideology.

## **Criterion Validity in Real Life - The Million Dollar Question**

This political test is a fairly simple linear relationship, and the criterion validity is easy to judge. For complex constructs, with many inter-related elements, evaluating the criterion related validity can be a much more difficult process.

Insurance companies have to measure a construct called 'overall health,' made up of lifestyle factors, socio-economic background, age, genetic predispositions and a whole range of other factors.

Maintaining high criterion related validity is difficult, with all of these factors, but getting it wrong can bankrupt the business.

## **Coca-Cola - The Cost of Neglecting Criterion Validity**

For market researchers, criterion validity is crucial, and can make or break a product. One famous example is when Coca-Cola decided to change the flavor of their trademark drink.

Diligently, they researched whether people liked the new flavor, performing taste tests and giving out questionnaires <sup>[5]</sup>. People loved the new flavor, so Coca-Cola rushed New Coke into production, where it was a titanic flop.

The mistake that Coke made was that they forgot about criterion validity, and omitted one important question from the survey.

People were not asked if they preferred the new flavor to the old, a failure to establish concurrent validity <sup>[3]</sup>.

The Old Coke, known to be popular, was the perfect benchmark, but it was never used. A simple blind taste test, asking people which flavor they preferred out of the two, would have saved Coca Cola millions of dollars.

Ultimately, the predictive validity <sup>[4]</sup> was also poor, because their good results did not correlate <sup>[6]</sup> with the poor sales. By then, it was too late!

---

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

#### **Links**

[1] [http://en.wikipedia.org/wiki/Criterion\\_validity](http://en.wikipedia.org/wiki/Criterion_validity)

[2] <https://explorable.com/pilot-study>

[3] <https://explorable.com/concurrent-validity>

[4] <https://explorable.com/predictive-validity>

[5] <https://explorable.com/survey-research-design>

[6] <https://explorable.com/statistical-correlation>