

Improving Literacy Brief

Understanding Screening

RELIABILITY

Reliability is a term that we, as educational professionals, frequently encounter, but just as often can take for granted. Simply, reliability is the consistency of a set of scores that are designed to measure the same thing. Suppose that a family is shopping at a supermarket and as the family makes their way to the produce section, the children decide to weigh a watermelon on five of the scales to figure out how much it costs. Reliability in measurement refers to how consistently the five scales provide the same weight for the watermelon.

WHY DOES RELIABILITY MATTER?

In this example, if each scale gave a different weight for the fruit, the shopper would not know the cost of the watermelon, and thus may not be able to make an informed decision about whether or not to buy it. The decision about how to act on a set of information requires that the information itself is trustworthy. When teachers, school psychologists, or other school personnel administer screeners of reading, there is typically an implicit trust or assumption that the obtained scores from the screener accurately reflect a student's ability, and there is little to no error in the score. In reality, reliability is a statistical property of scores that must be demonstrated rather than assumed.

FACTORS THAT INFLUENCE RELIABILITY

There are two broad factors that may impact the reliability of scores: *systematic errors* and *random errors*. Systematic errors include test-maker factors such as how items are constructed, errors that may occur in the administration of the assessment, and errors that may occur in the scoring of the assessment. Systematic errors may also include test-taker factors such as how tired the child is at the time of the assessment. Random errors tend to be more unpredictable in nature, such as the amount of noise in the classroom at the time of assessment. Test makers are charged with minimizing the types of systematic errors through the development and validation process of building screening assessments. Random errors cannot be controlled like systematic errors; however, statistical confidence intervals can be created to measure the uncertainty level of reliability for a set of scores. The wider the confidence interval, the greater the random error in reliability and the narrower the confidence interval the less random error in reliability.

FORMS OF RELIABILITY

At the outset of this brief, we defined reliability as the consistency of a set of scores that are designed to measure something. There are many forms or types of reliability. Internal consistency broadly refers to how well a set of item scores correlate with each other. Alternate form describes how well two different sets of items within an assessment correlate with each other. Test-retest is concerned with how stable two sets of scores are over a fixed period of time. Inter-rater is associated with how two different people observe a behavior and rate it in the same way. Each form of reliability are distinct and useful for their own purposes but should not be used interchangeably.

SUGGESTED CITATION

National Center on Improving Literacy. (2019). *Understanding Literacy Screening: Reliability*. <https://www.improvingliteracy.org/resource/understanding-literacy-screening-reliability>

WHERE DO I GO FROM HERE?

- For more information about identifying high quality screening tools: <https://intensiveintervention.org/>

RELIABILITY

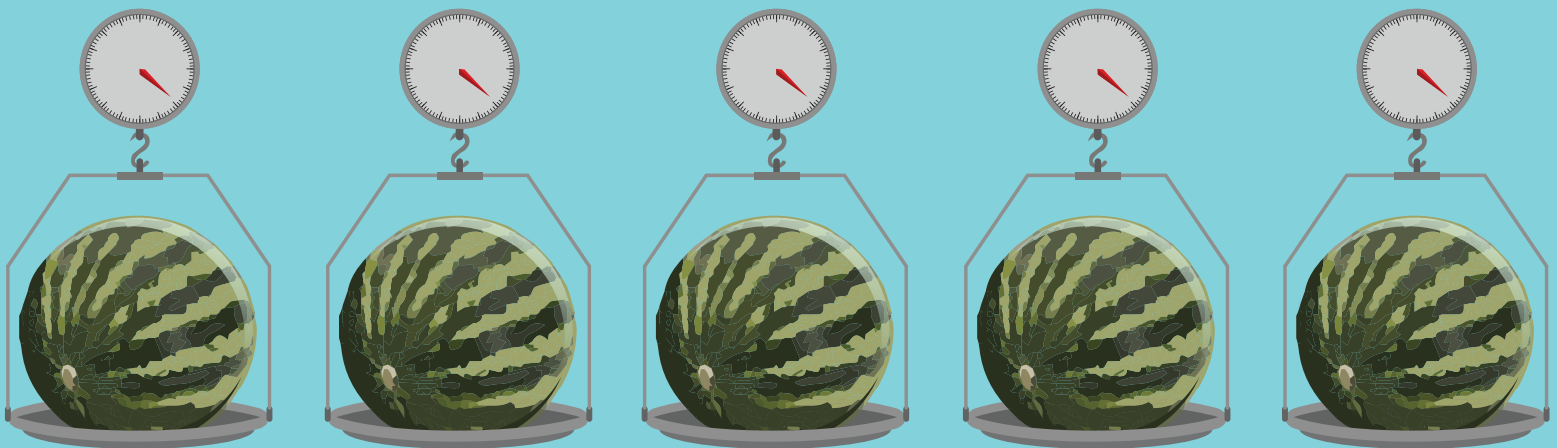
WHAT IS RELIABILITY?

Reliability is an important aspect of high-quality instruments used to measure academic progress. It is the consistency of a set of scores that are designed to measure the same skills.

Reliability is like using the produce scale at the supermarket

A child and his mother decide to weigh a watermelon on five different produce scales to figure out how much it costs. They want to know the reliability, or consistency, of the scales in providing the same weight for the watermelon.

When reading screeners are administered, there is typically an assumption that the scores are reliable, that they accurately reflect a student's ability and there is little to no error in the scores.



There are many kinds of reliability in reading screeners:

Internal consistency, or how well a set of item scores relate to each other.

Alternate form, or how well two different sets of items in a screener relate to each other.

Test-retest, or how stable two sets of scores are over a fixed period.

Inter-rater, or how two different people observe a behavior and rate it in the same way.