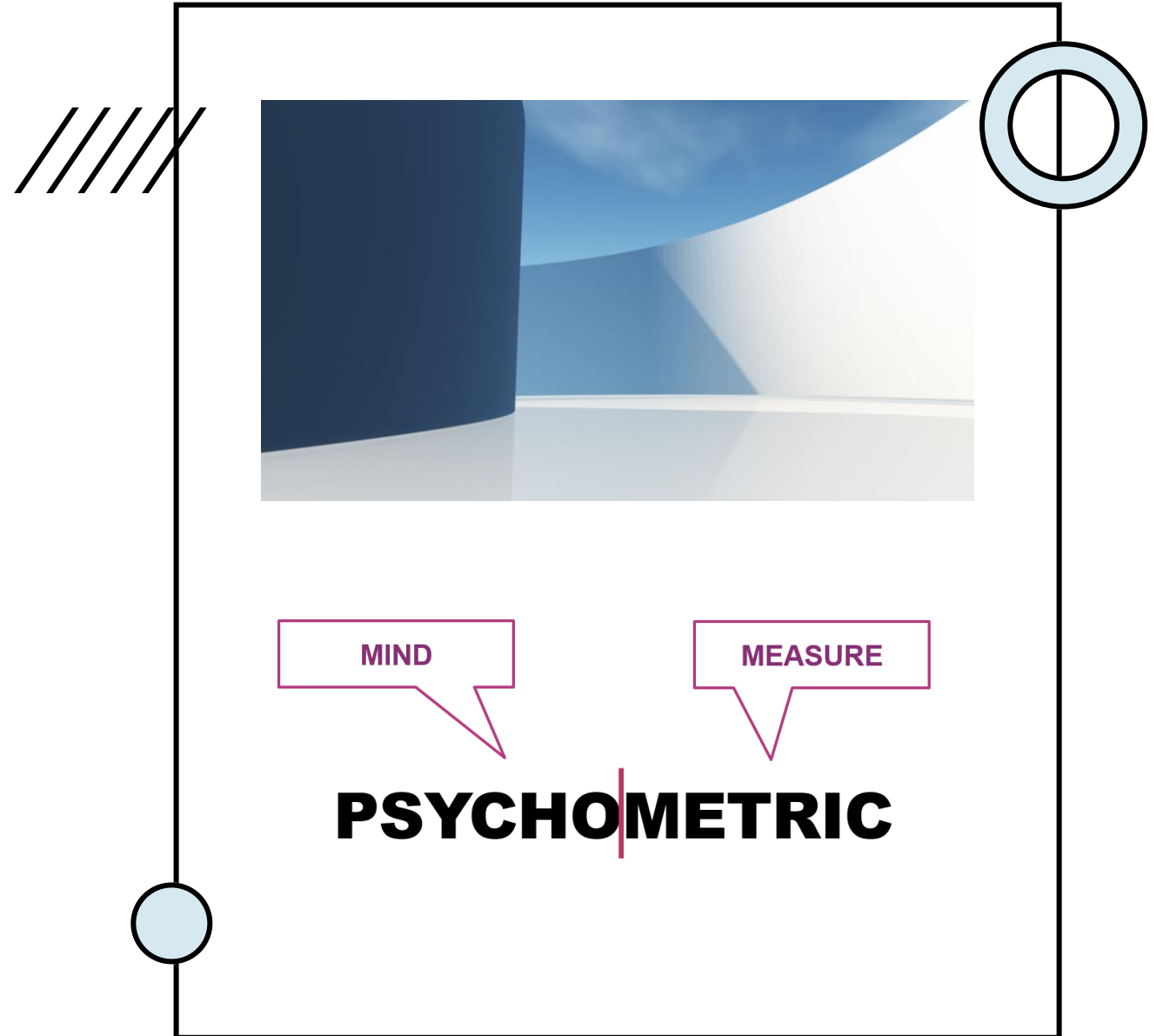


PSYCHOMETRIC PROPERTIES

Stacey Coleman



○ Psychometric Properties



The intrinsic components of a test are termed its psychometric properties



These are typical characteristics of tests that identify and define critical aspects of a test

E.g., its suitability or reliability for use in a certain context



Put simply, psychometric properties provide information about a tests adequacy, relevance, and validity



○ Psychometric Properties

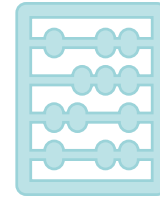
A good psychometric test must have three fundamental properties:



Psychometric
Reliability



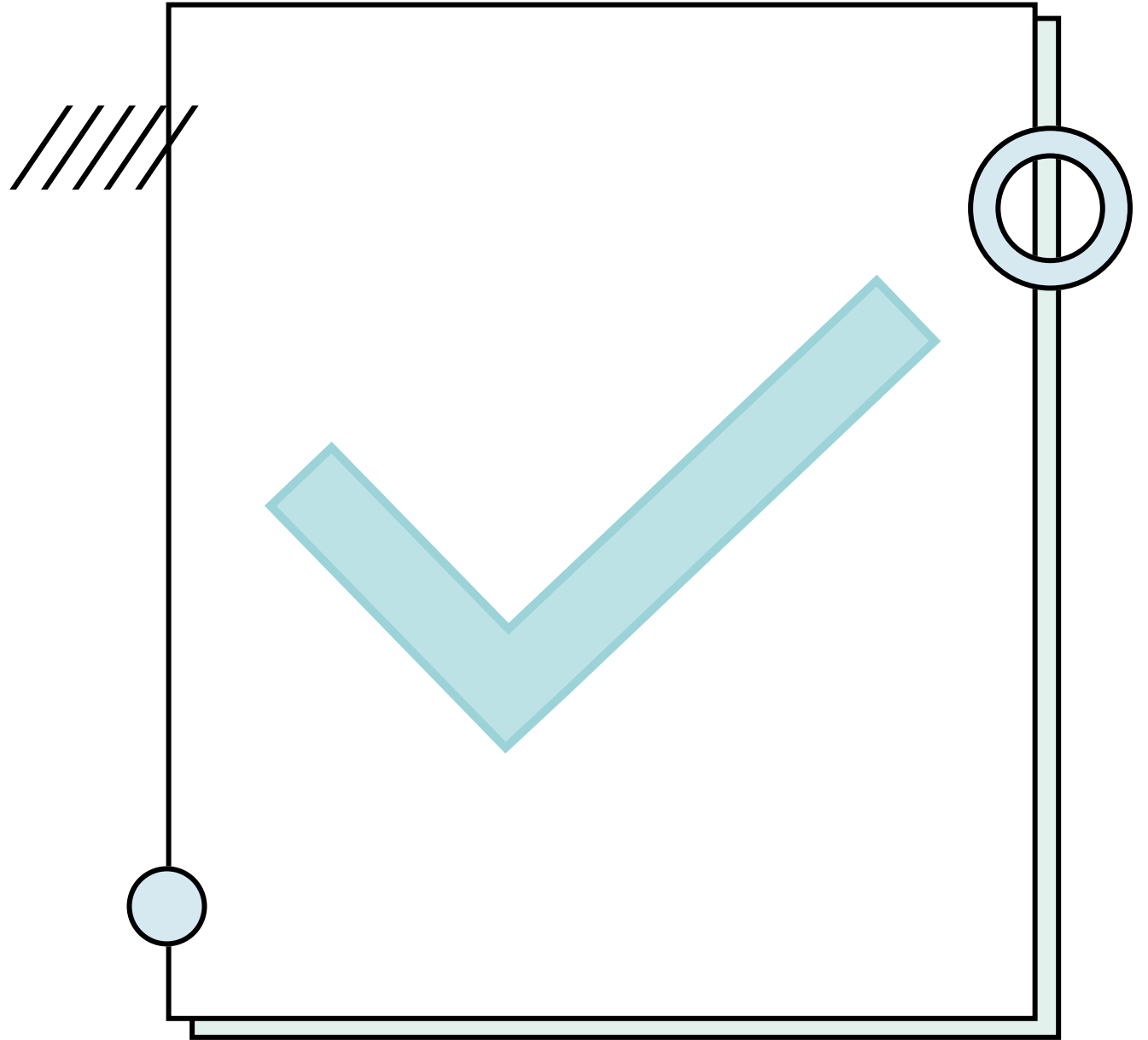
Psychometric
Validity



The Practice
of Norming



RELIABILITY



○ Reliability



Refers to the consistency or stability of a measure



The extent to which scores are accurate and without any measurement error



Any measure should yield a similar estimate on repeated use if the measured trait has not change



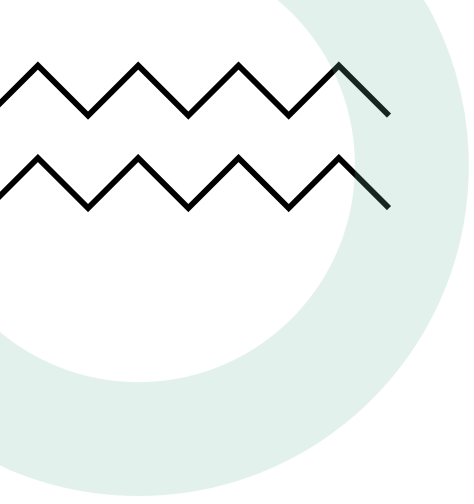
Three major types of reliability (important in P/ID)...





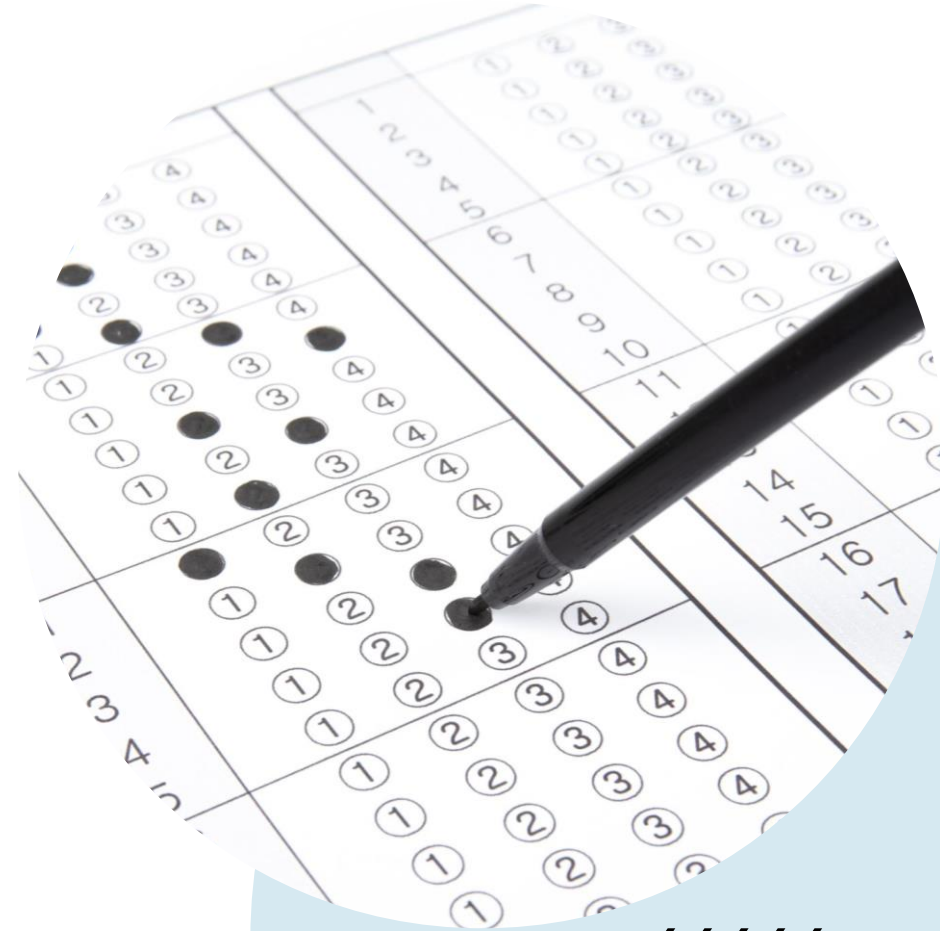
Reliability		
Test-retest reliability	Equivalent-form reliability	Internal consistency reliability

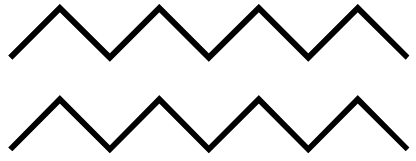




Test-Retest Reliability

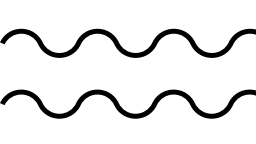
- Reveals stability of test upon repeated applications / over time
- Works by correlating two sets of scores
- *Coefficient of stability*





Equivalent-Form Reliability

- Reliability that reveals the equivalence of test scores between two versions/norms of a test
- Least popular in its use – difficult to create one test let alone two!
- *Coefficient of equivalence*



Internal Consistency



**Reliability that
reveals the
homogeneity of the
items within a test**



1. Split-half reliability

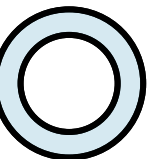
Divide items in half, correlate
between two halves

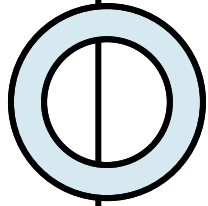


2. Cronbach's alpha

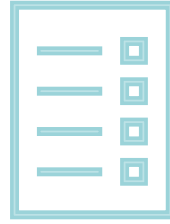
Response to each item
correlated with response to
every other item

Homogenous = high alpha





What can
have an
impact on
reliability
of
measures?



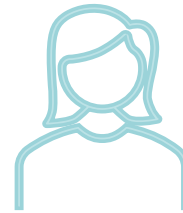
Construction of
items



Administration



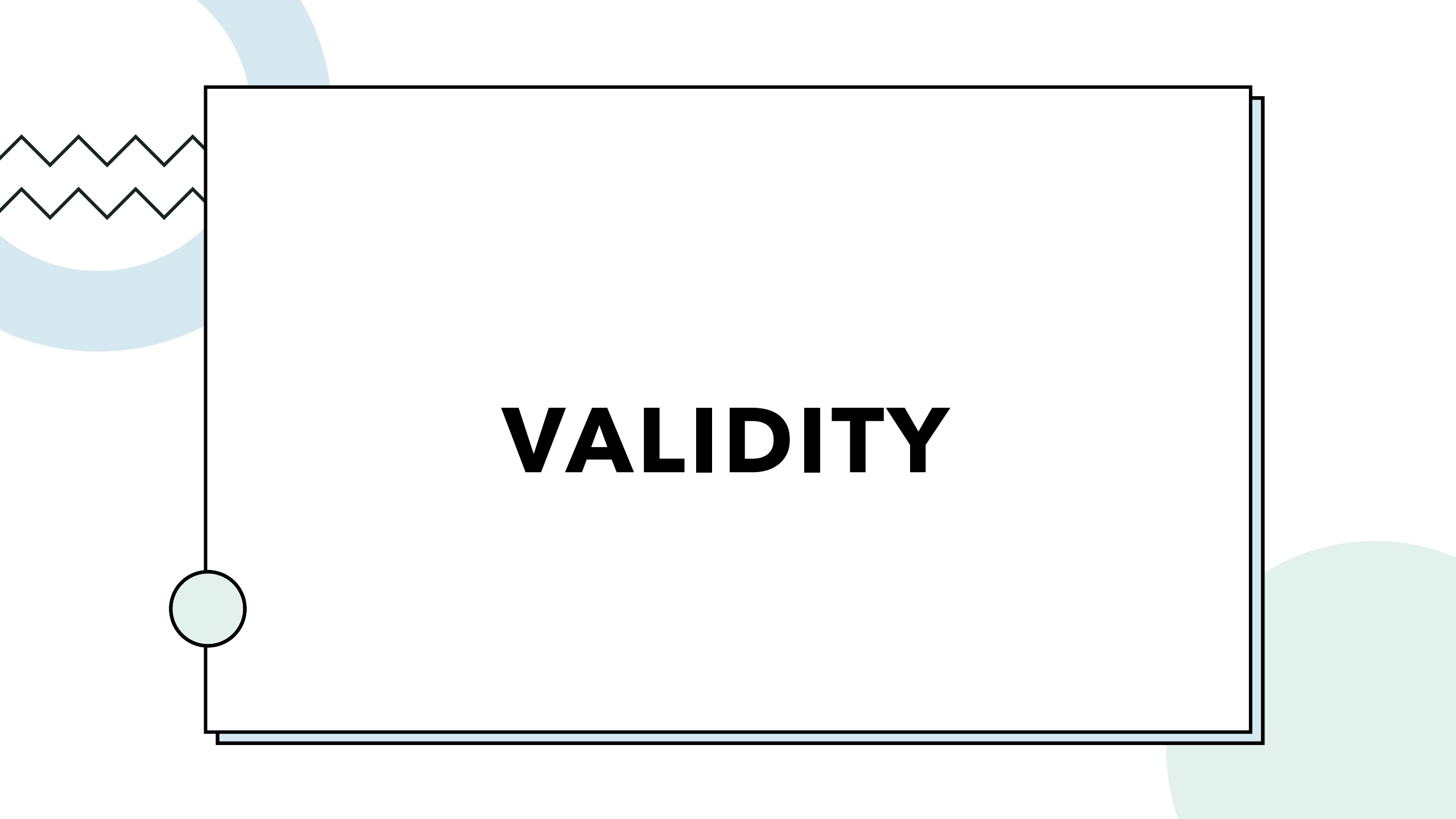
Environmental
factors



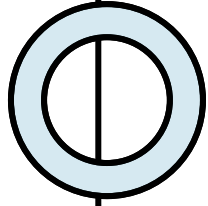
Test-taker



Scoring



VALIDITY



Validity



Refers to the accuracy and precision of a measure



The degree to which the test measures what it claims to measure



Validity refers to a tests appropriateness for predicting inferences about criteria



Validity

Construct
validity

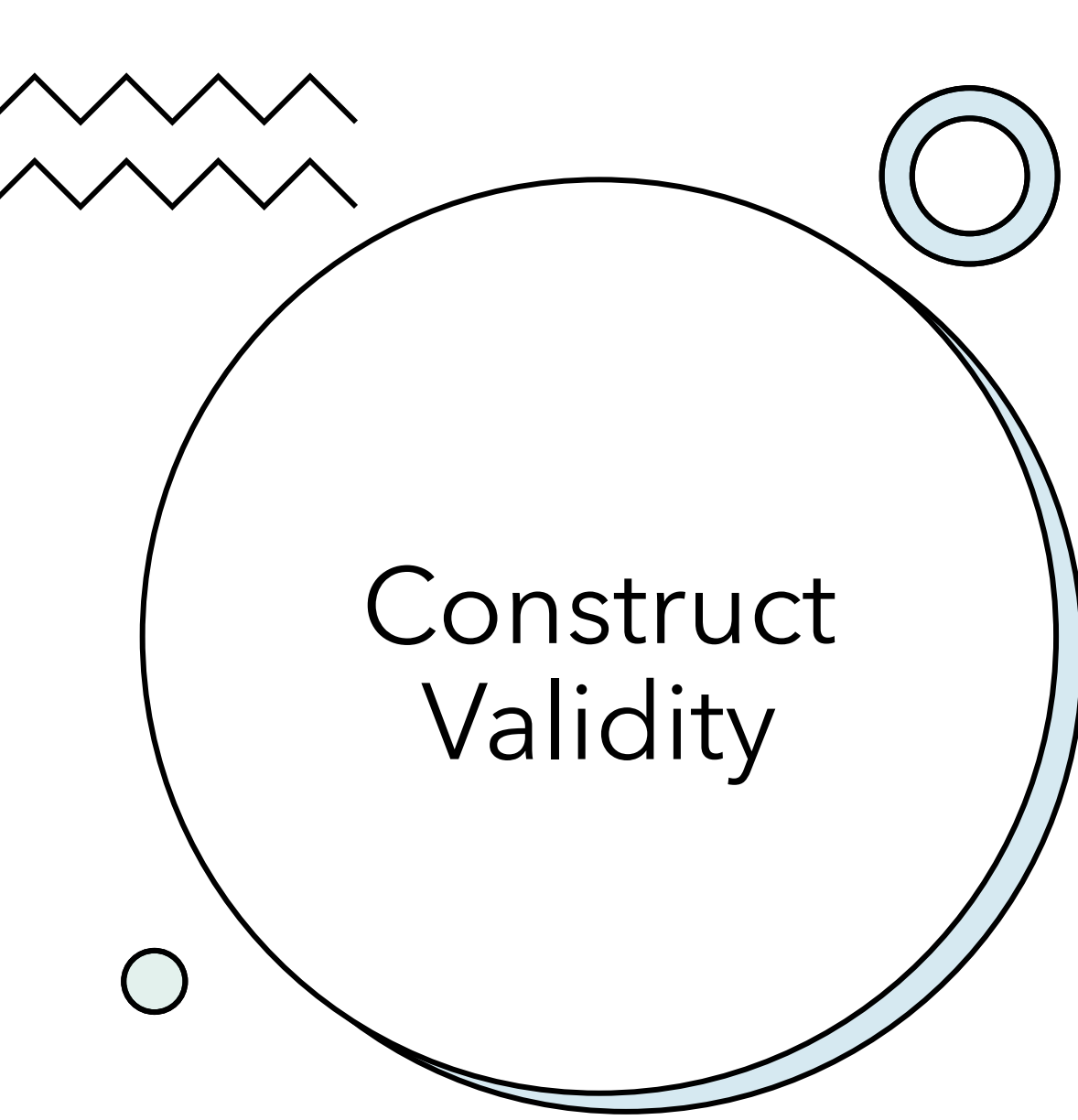
Criterion-related validity:

**Concurrent*

**Predictive*

Face validity





Construct Validity

- The degree to which a test is an accurate measure of the **construct** it purports to measure
 - Compare scores to other known measures of the same construct
 - We want the scores to be related to, or convergent with, other assessments, PLUS divergent with an assessment of an unrelated construct
- 

○ Criterion-Related Validity

Degree to which a test forecasts or is statistically related to a criterion



Concurrent

How well a predictor can predict a criterion
No time interval between collecting
predictor and criterion data



Predictive

Collect predictor information and use it to
forecast future criterion performance



○ Criterion-Related Validity



Expressed through a Validity Coefficient; a statistical index (correlation coefficient) that reveals the degree of association between two (or more) variables.



Often used in the context of prediction.



○ Face Validity



The appearance that items in a test are appropriate for the intended use

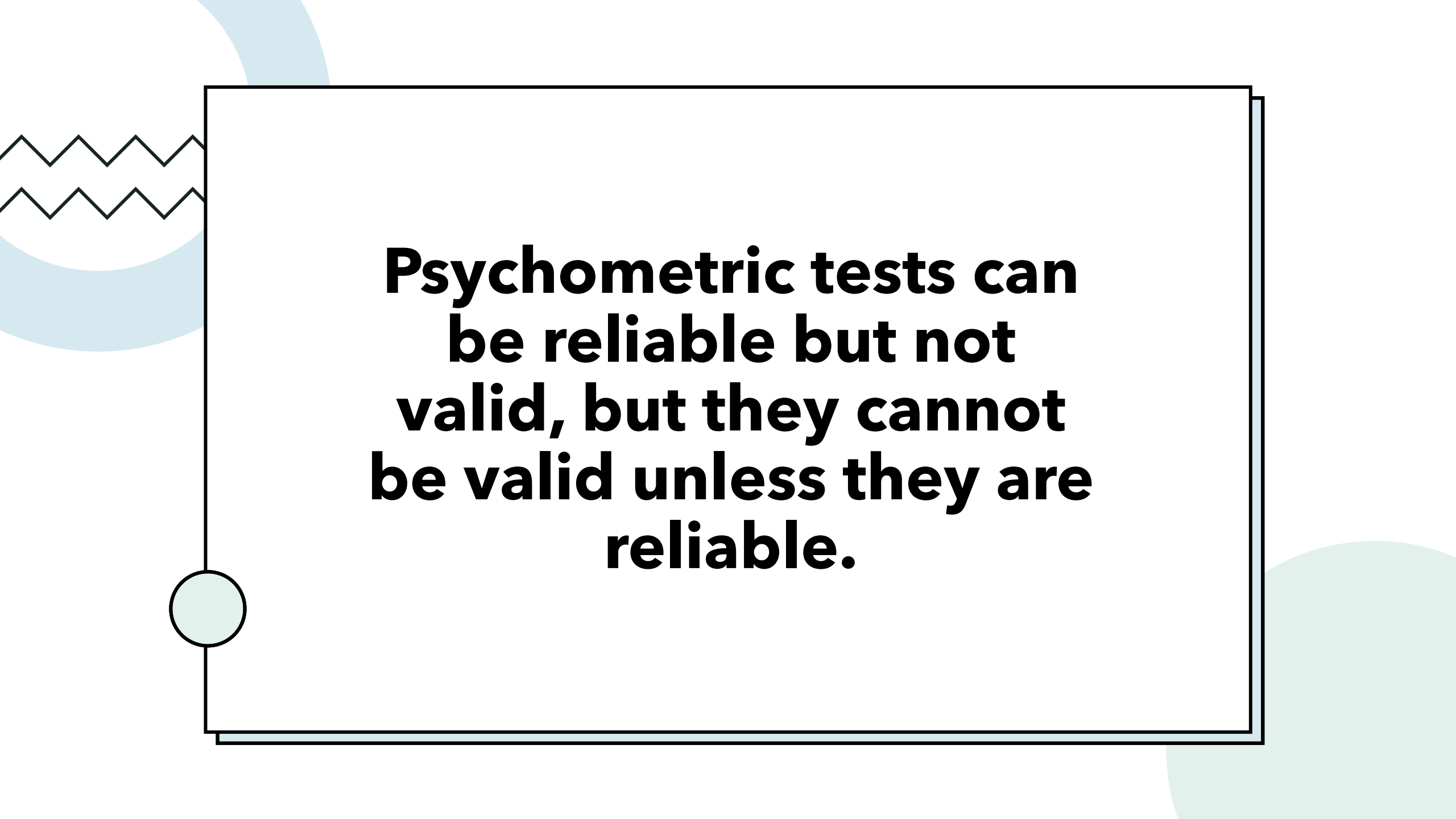


Judged solely on how the test appears 'at face value' to takers of the test

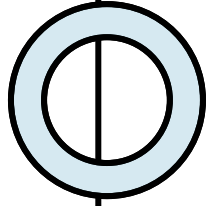


A very subjective assessment of validity





**Psychometric tests can
be reliable but not
valid, but they cannot
be valid unless they are
reliable.**

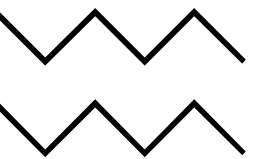


The importance of norming

Psychometric tests often normed against groups for comparison

Avoids evaluation of individual items and observes the total score of an individual against a representative sample

Considered important because score of 50 out of 75 on a psychometric test means nothing unless it is compared to performance of others at a similar level on the same test

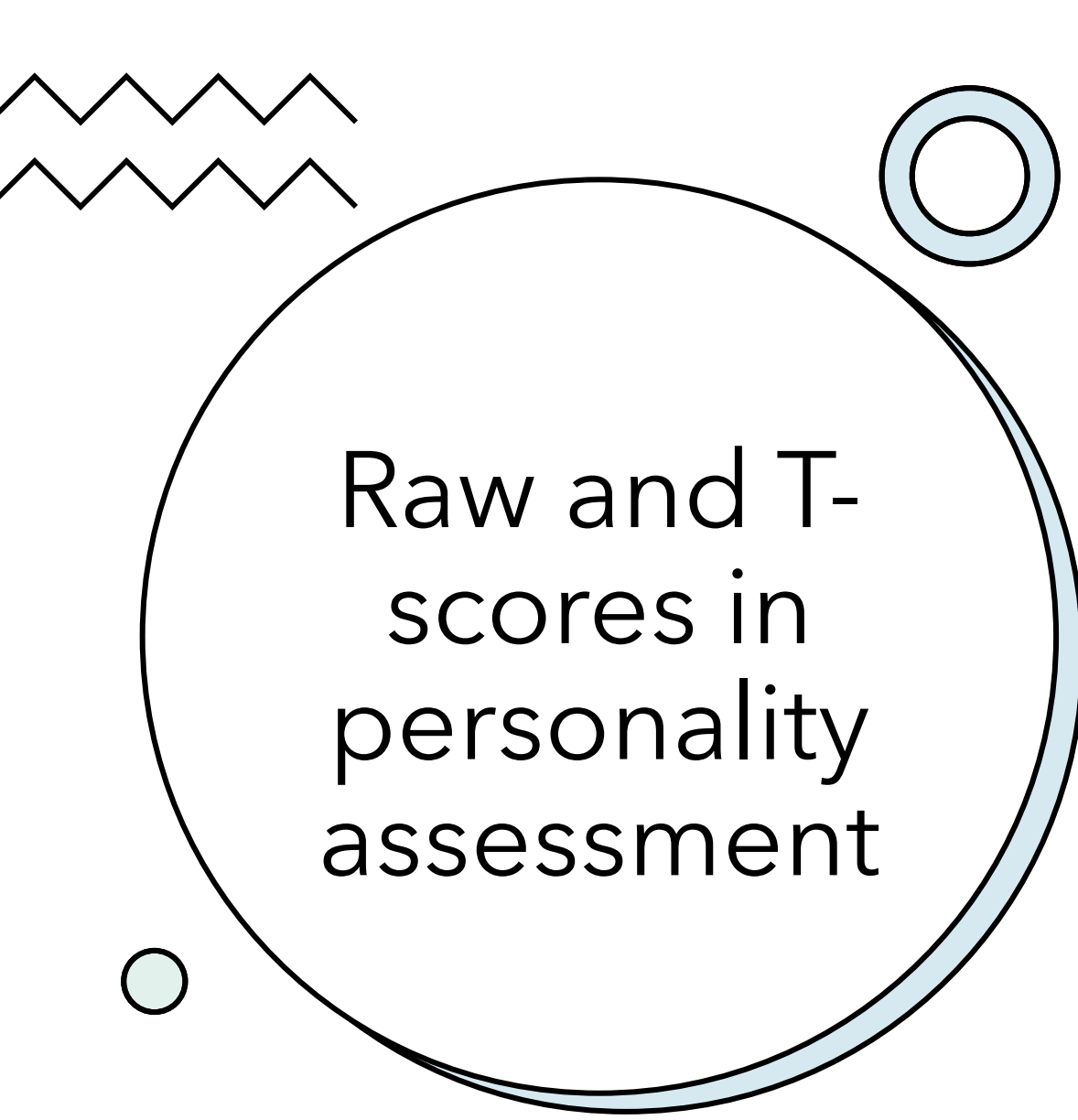


The importance of norming

Psychological constructs such as personality have no right/wrong answers, so cannot be 'marked' using percentages

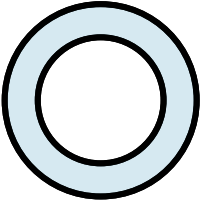
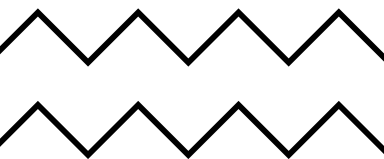
This is where norming becomes important to make sense of personality assessment scores



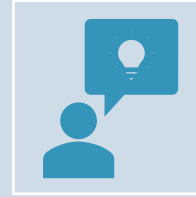


Raw and T-scores in personality assessment

- When we assess personality, the best practice is to compare results to a **norm group**. This allows us to determine how extraverted someone is **in comparison to other people** who are like them (according to a norm group of females, males, combined gender, etc)
- 



Raw and T-scores in personality assessment

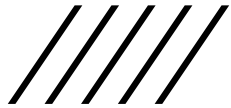


To be able to make this comparison, the results need to be presented in a meaningful way (psychometrics is often littered with technicalities and complex terminology, so we want to make it simpler!).



A t-score is a standardised way that scores can be presented to make them easier to understand.

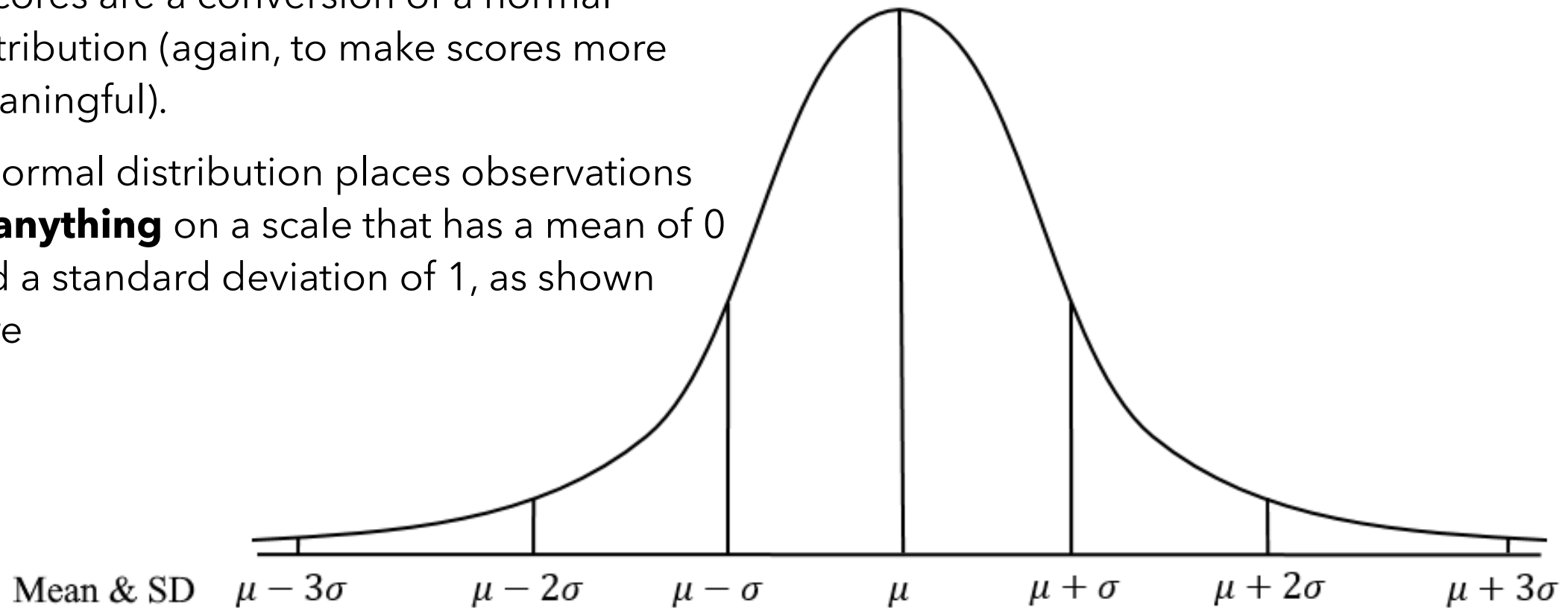
With the NEO assessments, for example, you would receive a raw score (the actual score as you reported your responses), and a t-score (the 'transformed', easier to understand score).





T-scores are a conversion of a normal distribution (again, to make scores more meaningful).

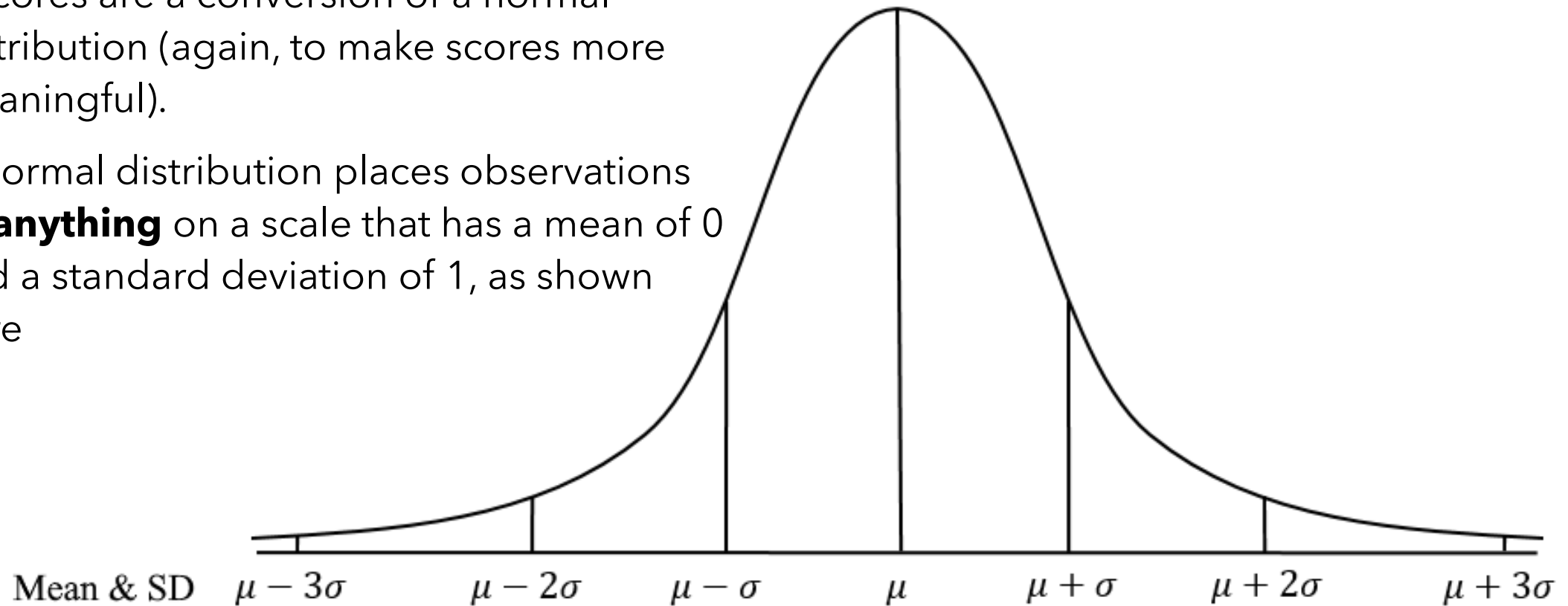
A normal distribution places observations **of anything** on a scale that has a mean of 0 and a standard deviation of 1, as shown here



○ T-scores and normal distribution

T-scores are a conversion of a normal distribution (again, to make scores more meaningful).

A normal distribution places observations **of anything** on a scale that has a mean of 0 and a standard deviation of 1, as shown here



○ T-scores and normal distribution

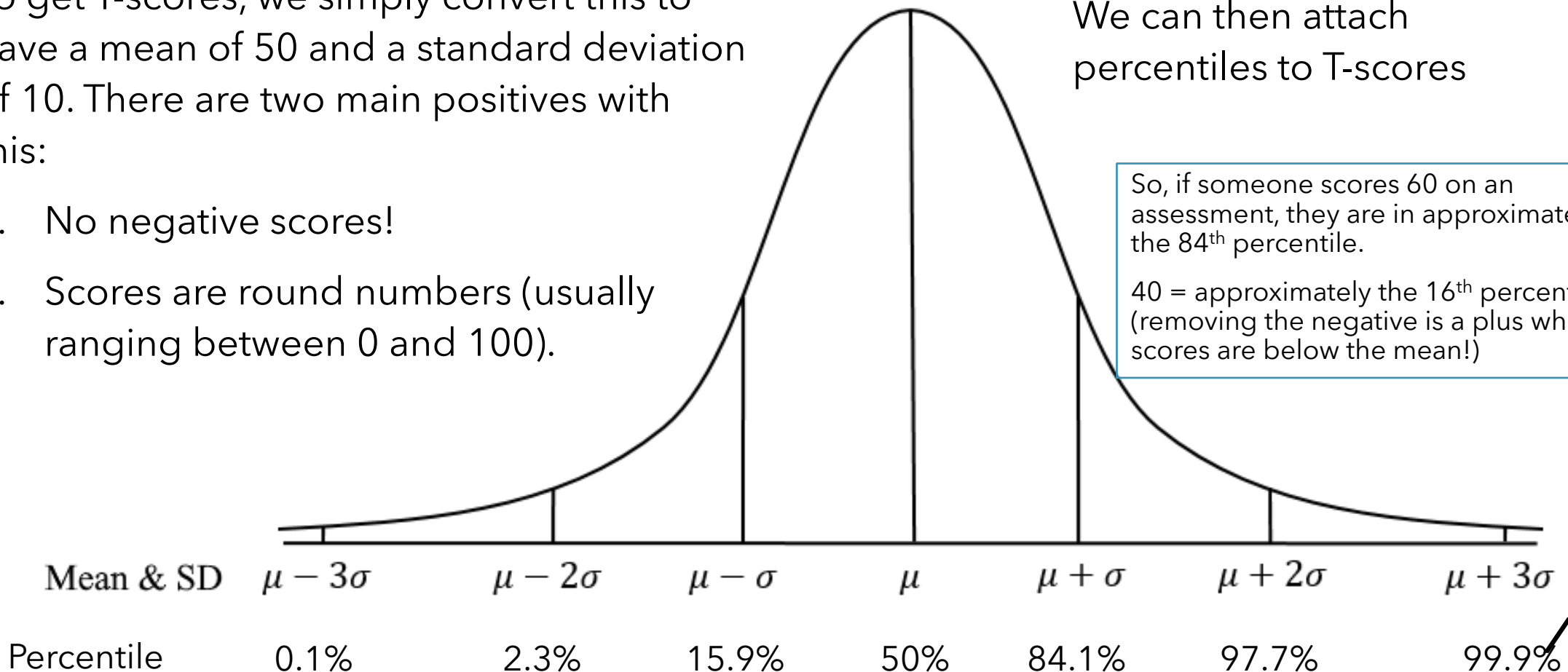
To get T-scores, we simply convert this to have a mean of 50 and a standard deviation of 10. There are two main positives with this:

1. No negative scores!
2. Scores are round numbers (usually ranging between 0 and 100).

We can then attach percentiles to T-scores

So, if someone scores 60 on an assessment, they are in approximately the 84th percentile.

40 = approximately the 16th percentile (removing the negative is a plus when scores are below the mean!)





Raw scores

The scores calculated based on participants responses to each question that makes up each facet

NEO-PI-3 Data Table

Scale	Raw Score	T Score	Range
Factors			
(E) Extraversion	--	44	Low
(C) Conscientiousness	--	68	Very High
(N) Neuroticism	--	47	Average
(O) Openness	--	45	Low
(A) Agreeableness	--	56	High
Extraversion Facets			
(E1) Warmth	18	41	Low
(E2) Gregariousness	12	37	Low
(E3) Assertiveness	20	58	High
(E4) Activity	18	51	Average
(E5) Excitement-Seeking	13	42	Low
(E6) Positive Emotions	19	51	Average

T-scores

Transformation of the raw scores to allow for comparison to a norm group

Range

How t-score relates to the average

There is **no new information** being provided with a t-score – it is simply a conversion or ‘repackaging’ of information that is easier to understand and explain

