

Reliability and Affecting Factors

By

Dr. Kshama Pandey

Associate Professor

Department of B.Ed./M.Ed.

M.J.P.Rohilkhand University,
Bareilly

Reliability

Reliability is the extent to which test scores are *not* affected by chance factors—by the luck of the draw. It is the extent to which the test taker's score does *not* depend on:

the specific day and time of the test (as compared with other possible days and times of testing),

the specific questions or problems that were on the edition of the test that the test taker took (as compared with those on other editions), and

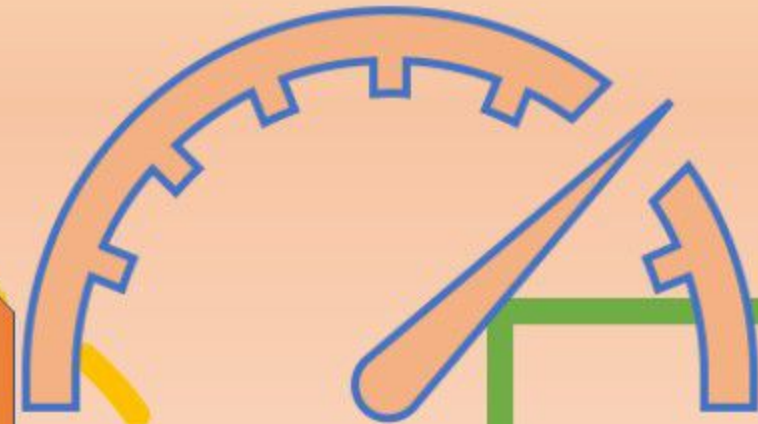
the specific raters who rated the test taker's responses (if the scoring process involved any judgment).

Reliability refers to the consistency of a measure and Psychologists consider three types of consistency:

Over time (test-retest reliability),

Across items (internal consistency),

Across different researchers (inter-rater reliability).



Types of consistency in Measuring reliability

Test-Retest Reliability

When researchers measure a construct that they assume to be consistent across time, then the scores they obtain should also be consistent across time.

Test-retest reliability is the extent to which this is actually the case. For example, intelligence is generally thought to be consistent across time.

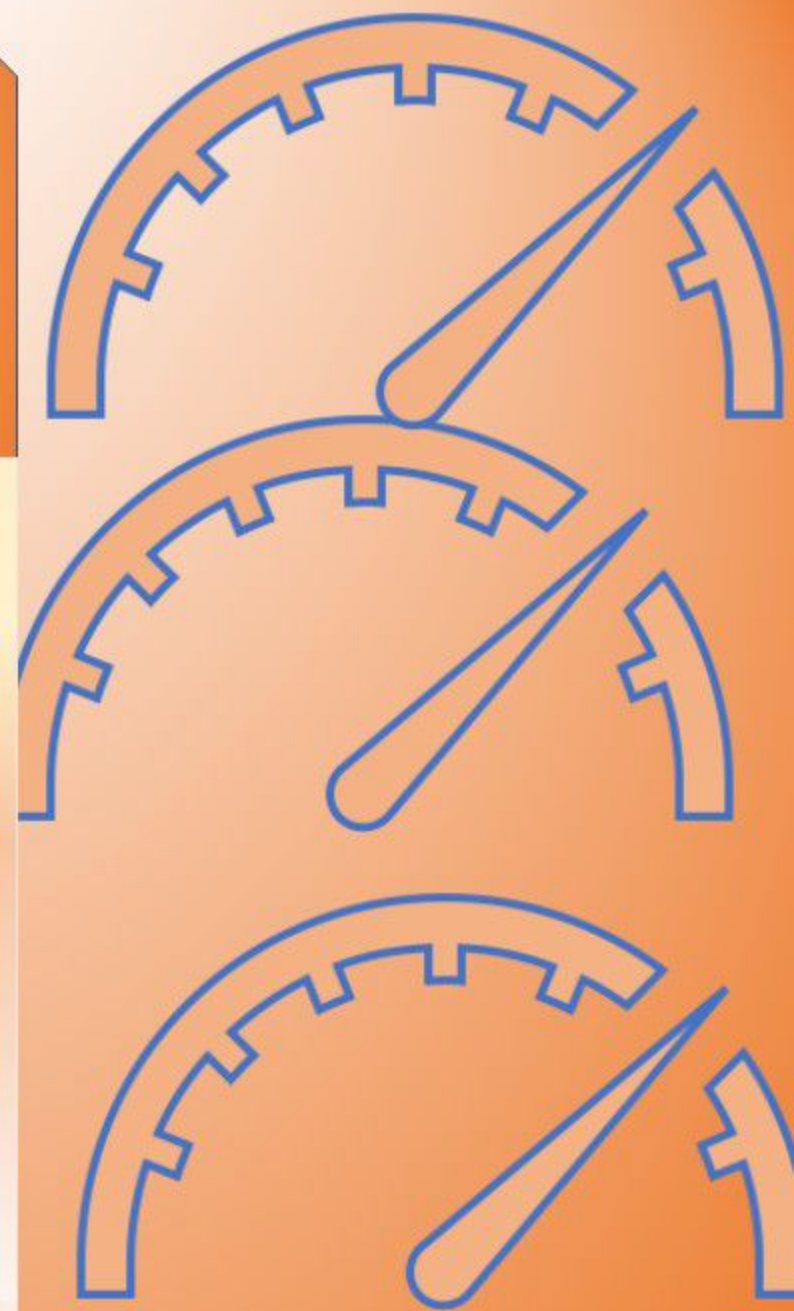


Test-Retest Reliability

A person who is highly intelligent today will be highly intelligent next week.

This means that any good measure of intelligence should produce roughly the same scores for this individual next week as it does today.

Clearly, a measure that produces highly inconsistent scores over time cannot be a very good measure of a construct that is supposed to be consistent.



Interrater Reliability

- Many behavioural measures involve significant judgment on the part of an observer or a rater.
- Inter-rater reliability is the extent to which different observers are consistent in their judgments.
- For example, if you were interested in measuring university students' social skills, you could make video recordings of them as they interacted with another student whom they are meeting for the first time.
- Then you could have two or more observers watch the videos and rate each student's level of social skills.





Interrater Reliability



To the extent that each participant does in fact have some level of social skills that can be detected by an attentive observer, different observers' ratings should be highly correlated with each other.


Inter-rater reliability would also have been measured in Bandura's Bobo doll study.

In this case, the observers' ratings of how many acts of aggression a particular child committed while playing with the Bobo doll should have been highly positively correlated.

Interrater reliability is often assessed using Cronbach's α when the judgments are quantitative or an analogous statistic called Cohen's κ when they are categorical.

Factors Influencing the Reliability of Test Scores

A) Intrinsic Factors:



- Length of the Test
- Homogeneity of Items
- Difficulty Value of Items
- Discriminative Value
- Test instructions
- Item selection
- Reliability of the scorer

B. Extrinsic Factors



- Group variability
- Guessing and chance errors
- Environmental conditions
- Momentary fluctuations

A) Intrinsic Factors: The principal intrinsic factors (i.e. those factors which lie within the test itself) which affect the reliability are:

Length of the Test

- Reliability has a definite relation with the length of the test .
- The more the number of items the test contains, the greater will be its reliability and vice-versa.
- However, it is difficult to ensure the maximum length of the test to ensure an appropriate value of reliability.

The number of times a test should be lengthened to get a desirable level of reliability is given by the formula:

$$n = \frac{r_{xx} (1 - r_1)}{r_1 (1 - r_{xx})}$$

Where r_{xx} = the desired reliability
 r_1 = the obtained reliability and
 n = number of times a test is to be lengthened.

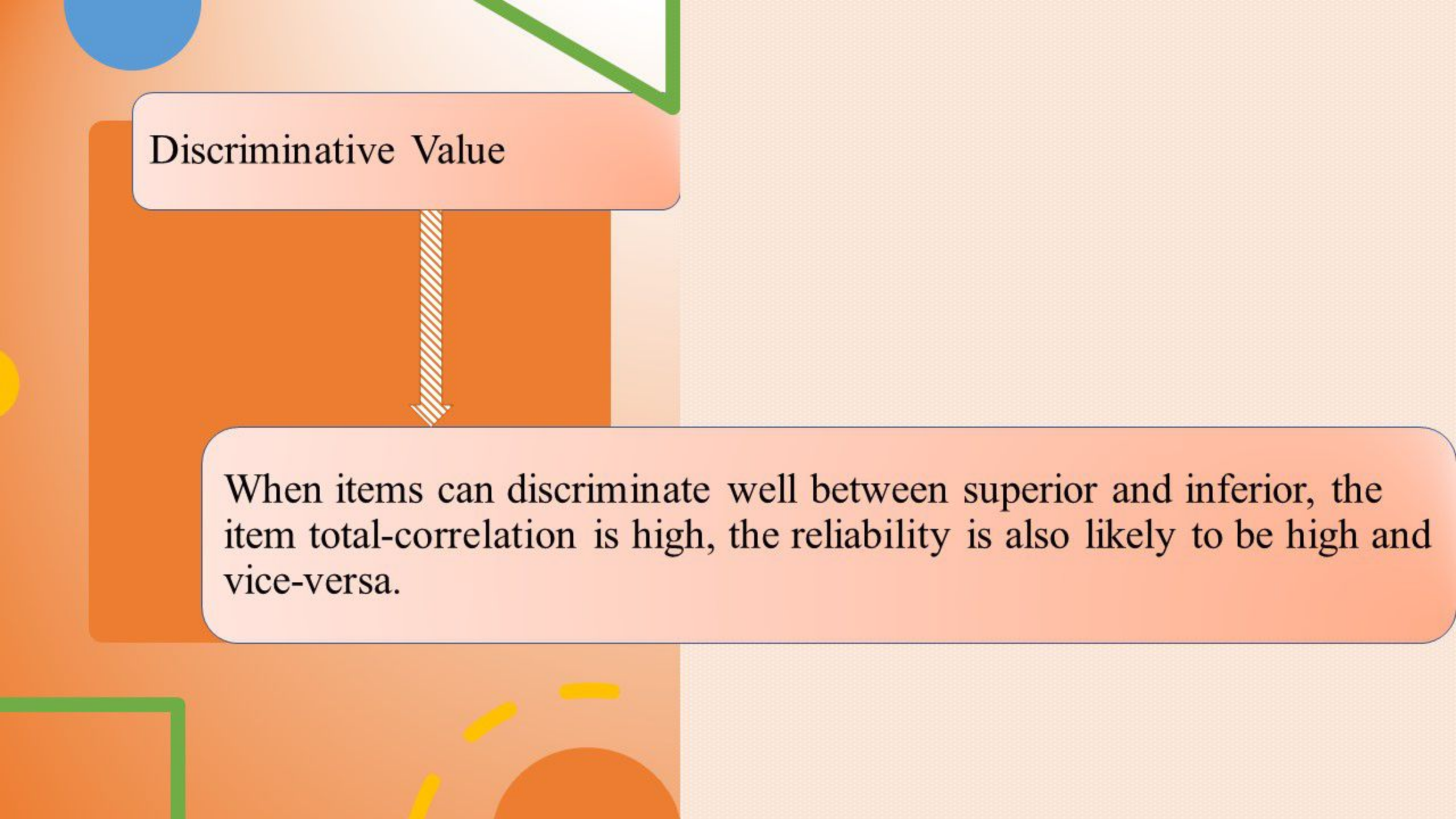
Homogeneity of Items

- Homogeneity of items has two aspects: Item reliability and the homogeneity of traits measured from one item to another.
- If the items measure different functions and the inter-correlations of items are 'zero' or near to it, then the reliability is 'zero' or very low and vice-versa.

Difficulty Value of Items

The difficulty level and clarity of expression of a test item also affect the reliability of test scores.

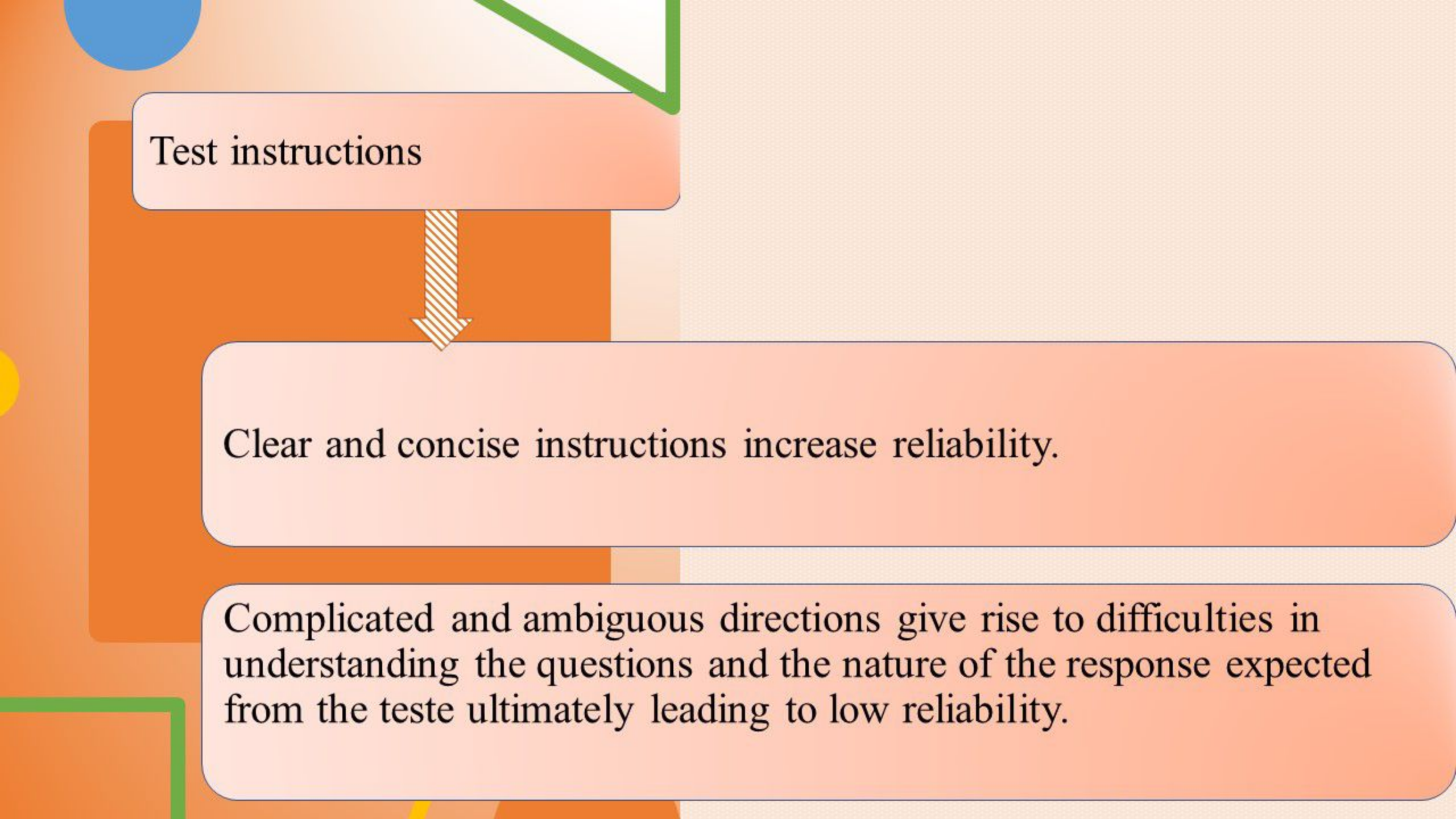
If the test items are too easy or too difficult for the group members, it will tend to produce scores of low reliability. Because both the tests have a restricted spread of scores.



Discriminative Value



When items can discriminate well between superior and inferior, the item total-correlation is high, the reliability is also likely to be high and vice-versa.




```
graph TD; A[Test instructions] --> B[Clear and concise instructions increase reliability.]; B --> C[Complicated and ambiguous directions give rise to difficulties in understanding the questions and the nature of the response expected from the teste ultimately leading to low reliability.];
```

Test instructions

Clear and concise instructions increase reliability.

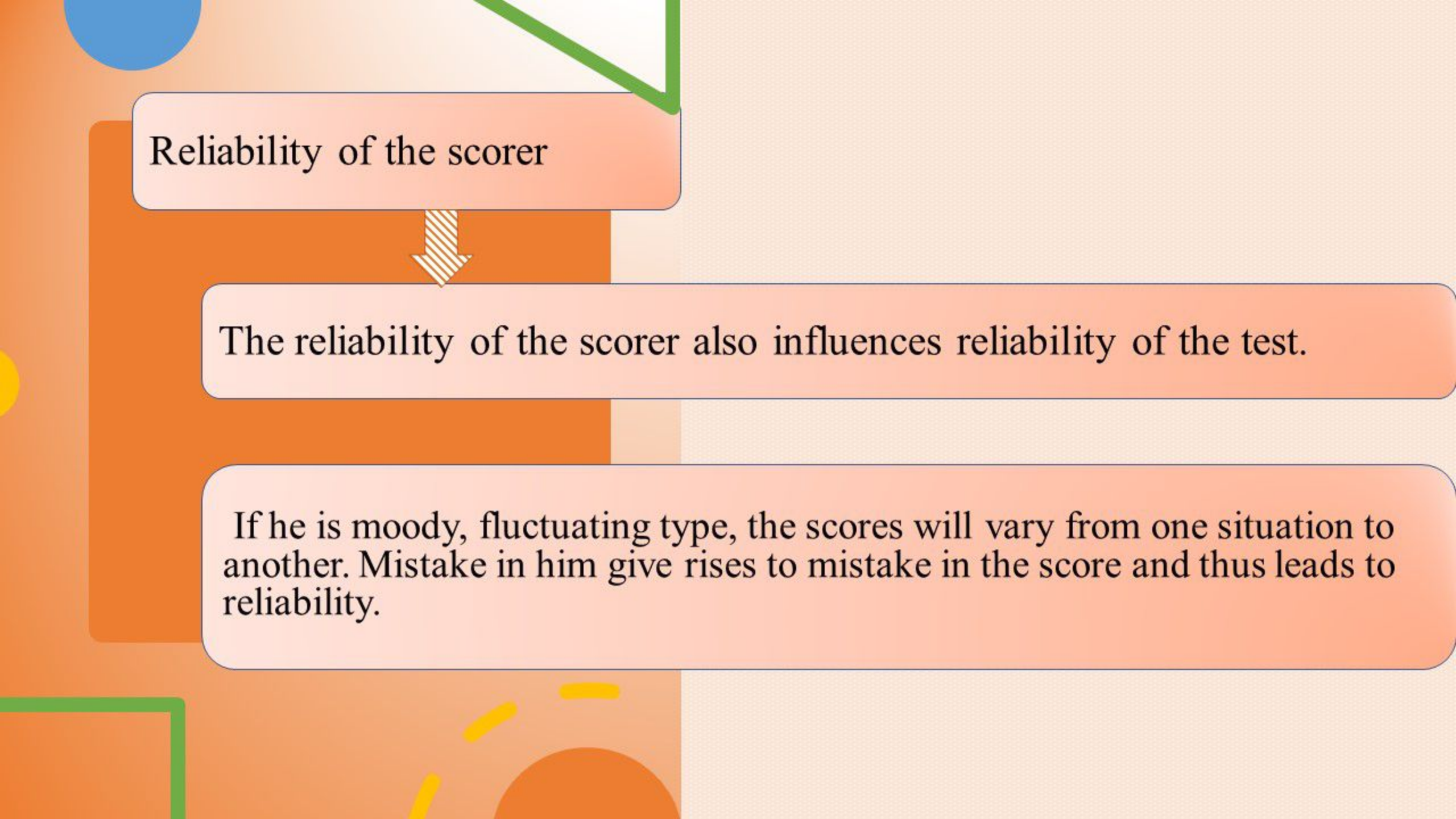
Complicated and ambiguous directions give rise to difficulties in understanding the questions and the nature of the response expected from the teste ultimately leading to low reliability.



Item selection

The diagram features a light orange rounded rectangle at the top containing the text 'Item selection'. A thick, downward-pointing arrow with a diagonal hatching pattern connects this box to a larger, light orange rounded rectangle below it. The background is a solid light orange color, decorated with abstract geometric shapes: a blue circle in the top-left corner, a green L-shaped line in the top-right corner, a yellow dashed line in the bottom-left corner, and an orange circle in the bottom-right corner.

If there are too many interdependent items in a test, the reliability is found to be low.



Reliability of the scorer

The diagram features a light orange background with abstract shapes: a blue circle in the top left, a green line in the top center, and a yellow circle in the bottom left. Three light orange rounded rectangular boxes are arranged vertically. The top box contains the text 'Reliability of the scorer'. A white arrow with diagonal stripes points from the bottom of this box to the middle box. The middle box contains the text 'The reliability of the scorer also influences reliability of the test.' The bottom box contains the text 'If he is moody, fluctuating type, the scores will vary from one situation to another. Mistake in him give rises to mistake in the score and thus leads to reliability.'

The reliability of the scorer also influences reliability of the test.

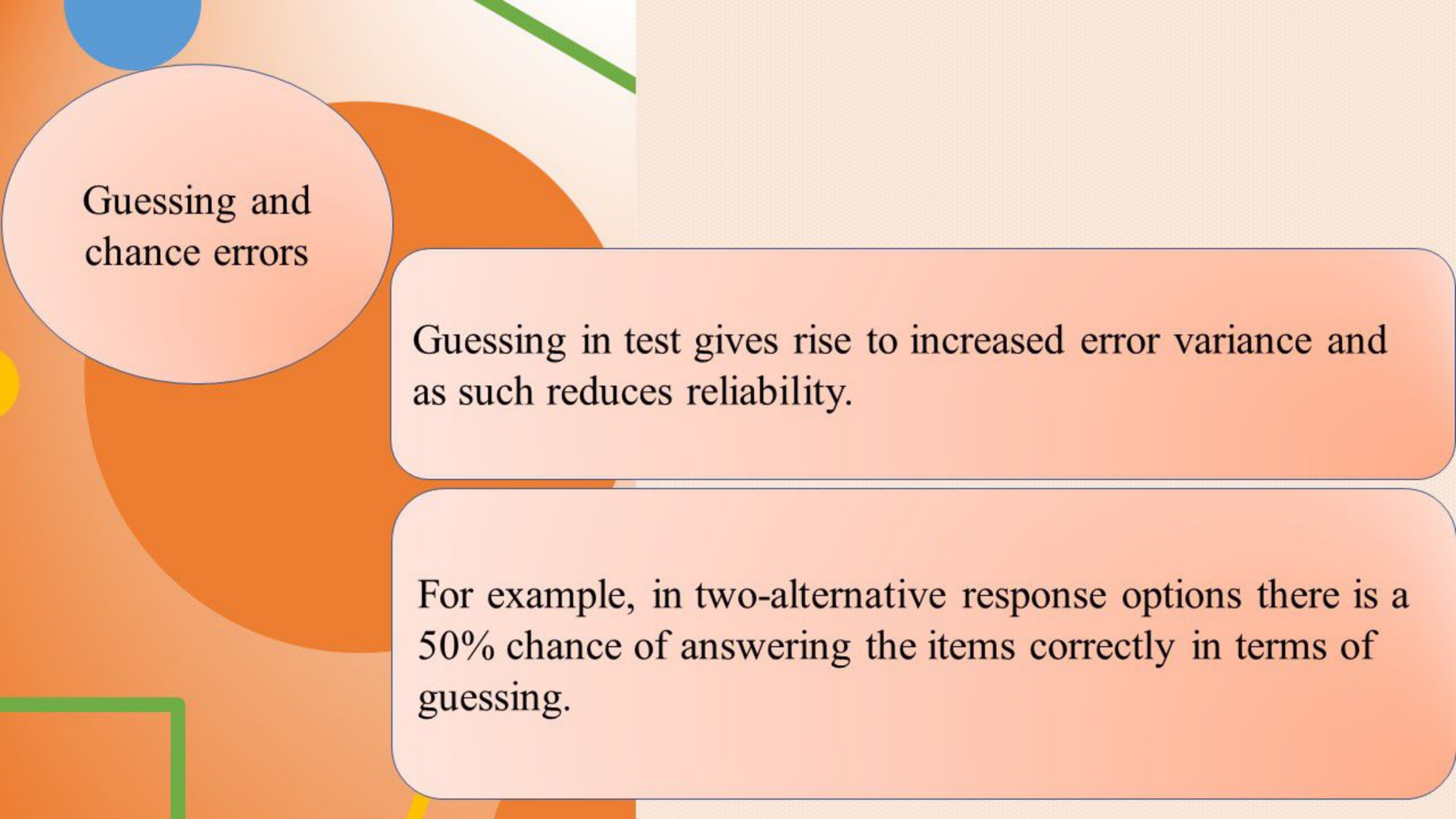
If he is moody, fluctuating type, the scores will vary from one situation to another. Mistake in him give rises to mistake in the score and thus leads to reliability.

Extrinsic Factors

The important extrinsic factors (i.e. the factors which remain outside the test itself) influencing the reliability are:

Group
variability

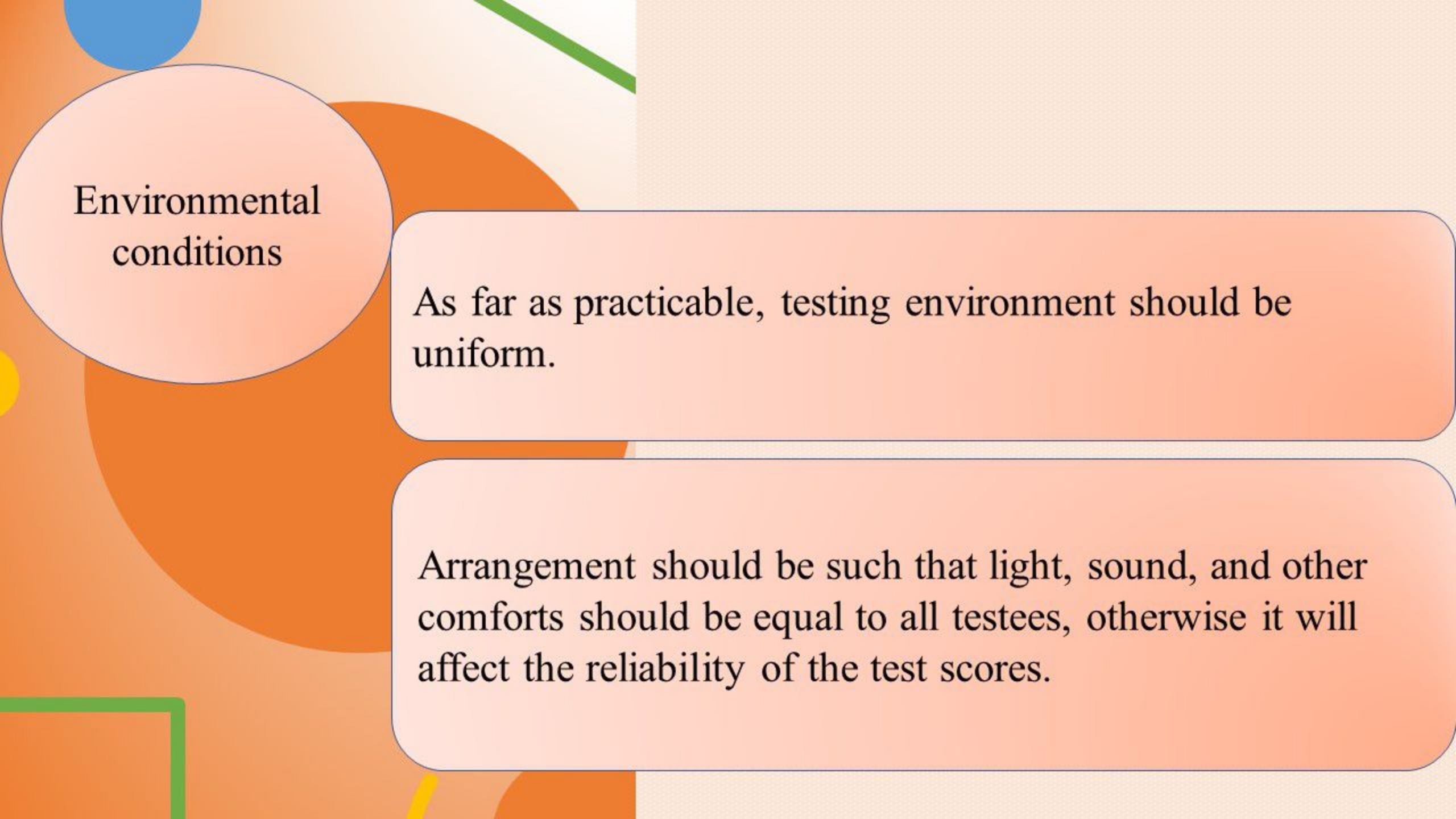
When the group of pupils being tested is homogeneous in ability, the reliability of the test scores is likely to be lowered and vice-versa.



Guessing and chance errors

Guessing in test gives rise to increased error variance and as such reduces reliability.

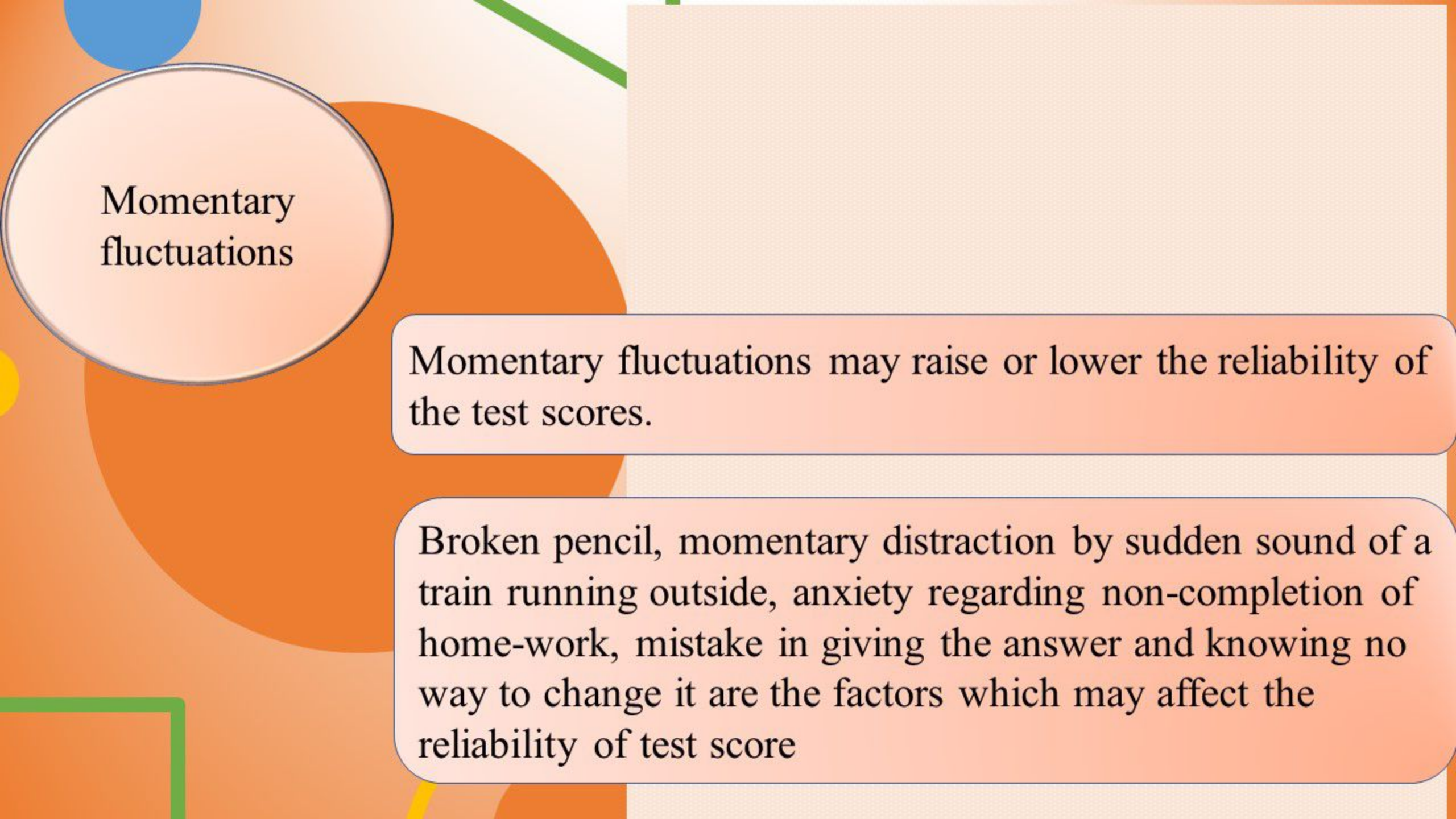
For example, in two-alternative response options there is a 50% chance of answering the items correctly in terms of guessing.



Environmental conditions

As far as practicable, testing environment should be uniform.

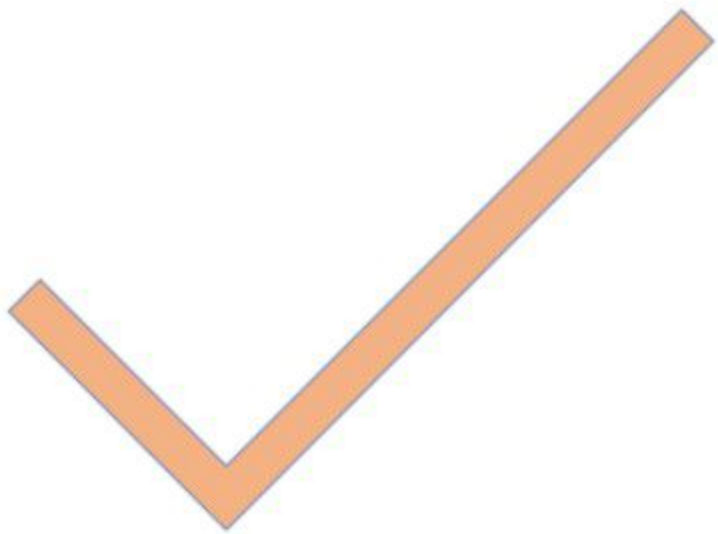
Arrangement should be such that light, sound, and other comforts should be equal to all testees, otherwise it will affect the reliability of the test scores.



Momentary fluctuations

Momentary fluctuations may raise or lower the reliability of the test scores.

Broken pencil, momentary distraction by sudden sound of a train running outside, anxiety regarding non-completion of home-work, mistake in giving the answer and knowing no way to change it are the factors which may affect the reliability of test score



THANK YOU

