

Standardization usually includes three elements.

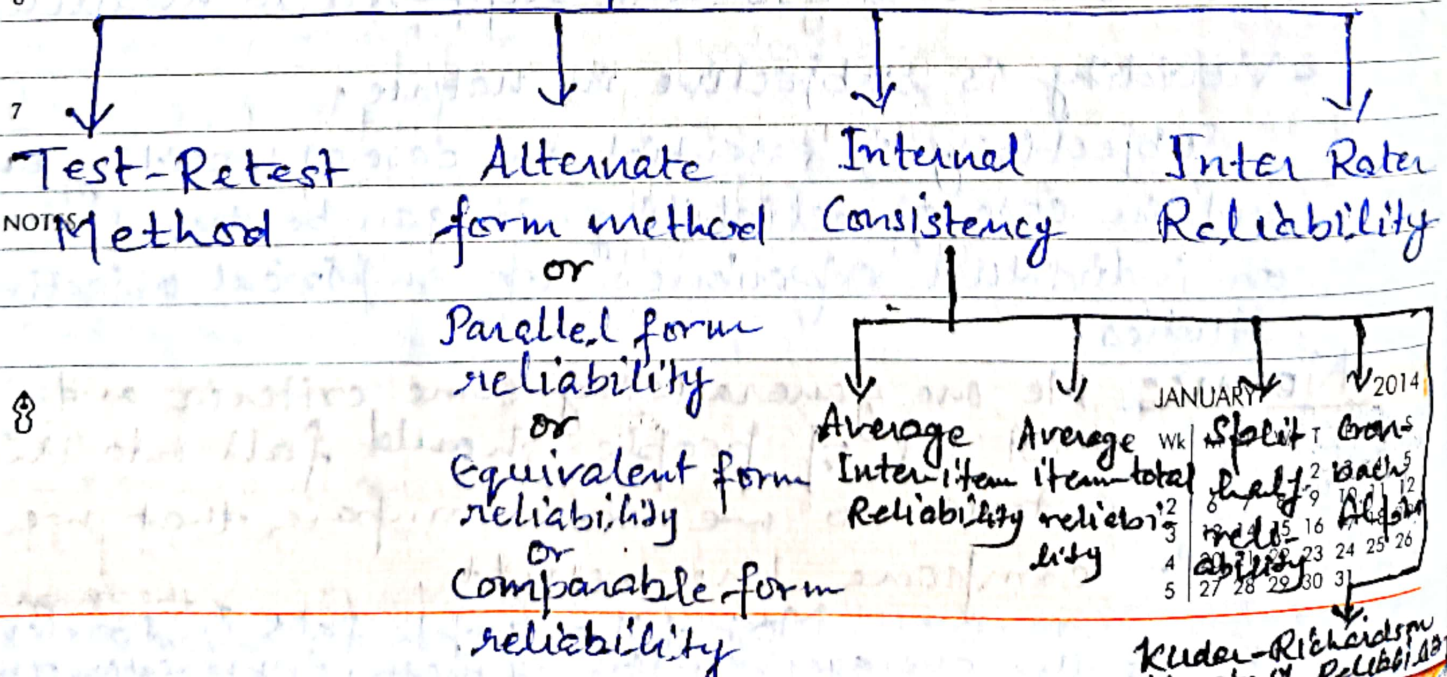
Reliability → able to repeat the test once and again and get consistently high score that means the reliability is good or high.

Validity → The test is to measure what it should be suppose to measure.

Norms → Generalizing those set of test and able to compare group of people based on norms which are established

There are various measures to understand the reliability. Four basic form of classification that we talk about reliability. —

Types Of Reliability



Test-Retest Reliability:

If you are doing a test and retesting it after a given time, it would be a test-retest reliability. For ex. — I am doing a test on 10 AM today and I am doing the same test on 5 PM and comparing the result of these two tests. This comparison can be done by simple correlation or Pearson's correlations and if there is a high correlation I can say that the test are reliable or reliability is high. A reliability of around 0.80 and more is considered a good reliability score for a test should be considered as a standardized test.

Ex:-

Student-ID		Exam Score	(after 6 months) Repeat Score (Same)
Length of the test remains the same	1	45	49
	2	58	60
	3	32	41
	4	54	58
	5	48	54

Here, consistency is important in the score of students. I can't say the student from 48 is directly jumping to 98. So there is a kind of consistency that is maintained the score of the student.

FEBRUARY 2014

Wk	M	T	W	T	F	S	S
5						1	2
6	3	4	5	6	7	8	9
7	10	11	12	13	14	15	16
8	17	18	19	20	21	22	23
9	24	25	26	27	28		

Limitations of this approach:-

- (i) Memory effect/carry over effect
- (ii) Practice effect
- (iii) Absence

The assumption behind this approach is that there will be no substantial changes in the measurement of the construct in question, upon administration on separate occasions. The time gap that is given between measure is of critical value, the shorter the time gap, higher the correlation value and vice versa.

12 Parallel or Alternate Form Reliability?

Parallel forms reliability compares two equivalent form of a test that measures the same attributes. The two forms use different items.

This alternate form reliability is also known as parallel reliability because you are conducting test in two forms of a same point of time.

For Ex:— A person is appearing in grade 10 examination. He is asked to answer 10 one line questions and also 10 essay type questions. Now essay type question requires descriptive knowledge but one line question you need to be very sure about the correct answer. So it's a kind of much more objective question you must be very sure you are writing about. At the same point of time where you are appearing for 10 grade science paper, you are being judged objective skills as well as descriptive skills. This is known as parallel form of reliability.

alternate form delayed
↓
time sampling

NOTES

alternate form & immediate
↓
content sampling

JANUARY 2014						
Wk	M	T	W	T	F	S
1			1	2	3	4
2	6	7	8	9	10	11
3	13	14	15	16	17	18
4	20	21	22	23	24	25
5	27	28	29	30	31	

- ✓ test given on different time — error associated with time sampling.
 ✓ test given on same day — only source of variation are random error and the difference between the forms of the test

WEDNESDAY
 DAY 008-357

So we are trying to judge two kinds of reliability at a same time the objective characteristics as well as descriptive characteristics. It can be vary on time as well as form but it can be same time duration or different time duration. Here we are judging pre-test and post-test result as well. Pearson Product moment correlation coefficient is used for an estimate of the reliability.

Student ID	Exam A Scores	Exam B Scores
1	57	58
2	Pre test 54	Post-test 45
3	Type A 41	45
4	Type B 32	41
5	Time 1 Type A 58	Time 2 Type B 57

In this case 5 students who are appearing for the exam, they are being given exam A and exam B and they have a score which are consistent like for the 5th student there is a decline in the marks but is not huge. So there is a kind of consistency here and this consistency is maintained throughout the 5 students appearing for this examination.

NOTES

So you can have for alternate form, I can

say there can be two forms which are independent for each other one is MCQ other is essay type question or one is one line question and other is descriptive question.

FEBRUARY 2014

Wk	M	T	W	T	F	S	S
5						1	2
6	3	4	5	6	7	8	9
7	10	11	12	13	14	15	16
8	17	18	19	20	21	22	23
9	24	25	26	27	28		

It can be varied on time as well as on form. So it is definitely varied on form but that can be same time duration and diff time duration.

Inter Rater Reliability Or Inter Observer Reliability or Inter Coder Reliability :

- ↳ Consensus in observations among different people (assuming observation are the same).
- ↳ Degree of scoring stability a measure yields across multiple observers.
- ↳ Measure of agreement among raters
- ↳ Improving Rater Reliability — Training/ Guided Revisions.

What affects Inter-Rater Reliability?

- ↳ Halo effect (One dimension influencing other)
- ↳ Stereotyping (Impressions on group may influence individual)
- ↳ Perception differences (rater viewpoints and past experiences)
 - ↳ Difference based on observation
- ↳ ~~List~~ Leniency/ Stringency Error (Lack of knowledge)

NOTES

JANUARY

Wk	M	T	W	T	F	S	S
1			1	2	3	4	5
2	6	7	8	9	10	11	12
3	13	14	15	16	17	18	19
4	20	21	22	23	24	25	26
5	27	28	29	30	31		