Instrumentation

Dr. Sangob Laksana
Dean, Graduate School of Education

Instrumentation is the description of how to measure a research variable. It is the specification of the instrument or the technique to be used for obtaining the quantitative data or qualitative data representing the value or the characteristics of the variable being studied.

The Writing of "Instrumentation" in Chapter 3

Under the topic *Instrumentation* in the research proposal, the following aspects should be clearly identified:

1. **The name of the instrument**.

   For example:

   - Observation
   - Test
   - Questionnaire
   - Interview
   - Review of documents and evidence

2. **The variable(s) to be measured by this instrument.**
3. **Development of instrument.**
4. **How to determine content validity**
5. **How to conduct the tryout and how to determine the reliability**
6. **The structure of items and variable to be measured**
7. **The criteria of interpretation**

The structure of instrument is the identification of item numbers to be used for measuring a variable.

Example, The questionnaire consists of the following parts:

Part I: The checklist item 1 – 10 measuring the demographic variables

Part II: The rating scale item1 – 20 measuring the teacher perception on school management

Part III: The open-ended item 1 – 3 asking about comments, observation, and recommendation

Development of Instrument

The instrument development comprises the following steps:

1. **Provide the operational definition of variable** in terms of observable and measurable aspects. Theoretical support is in need.

For example,

Job Satisfaction *is the individual perception of the extent to which the job meets his/her needs and expectation.*

2. **Decide on the nature of variable characteristics**

- **Cognitive Ability**, e.g. knowledge, comprehension, application, analysis, synthesis, evaluation, critical thinking, creativity, intelligence, skills, ability
- **Affective Characteristics**, e.g. attitudes, interest, need, satisfaction, personality traits
- **Psychomotor Skills or Performance**, e.g. manual/practical skills in arts, music, sports; use of equipment, tasks performance
- **Personal Attributes** like perception, opinion, demographic variable (*e.g. education level, income, age, nationality***)**

3. Decide on the mode of measurement

1) **Test** is the measurement of individual's best or maximum performance, with the criterion correct or incorrect response. Mostly the test is used for assessing the individual cognitive ability. The types of test are as followed:

- Objective test (True-false, Matching, Multiple-choice)
- Short answer test
- Essay test
- Oral test
- Performance test or Authentic Test (Test of Process, and/or Product)

2) **Non-Test Instrument**, the instrument that gather the individual data or information with out predetermined right/wrong response. Mostly it is designed to assess qualitative or quantitative aspects of variables aside from cognitve abilities. The non-test is frequently used for assessing *interest, need, attitude, perception, motivation, values, practice or performance, and personality traits.*

The instrument and the assessment techniques under the non-test are as follows:

2.1 Observation, emphasis on process and behavior

2.2 Interview

2.3 Inventories or Scales, e.g. attitude scales/inventories

2.4 Questionnaire/Opinionnaire. When respondents answer questions or respond to statements in writing, emphasis on factual information, it is the **questionnaire**. When opinions rather than facts are desired, an **opinionnaire** is used. These two purposes can be combined into one form that is usually referred to as a ***questionnaire.***

2.5 Evidence study or document analysis, e.g. review the school report, review the lesson plan, achievement record, etc.

4. Decide on the recording mode, the record of the measurement data:

1. Description
2. Score of correct answer
3. Checklist
4. Rating scale

5. Develop the measuring items guided by the operational definition of the variables and based on the principle of item writing quality, or use the standardized instrument

6. **Evaluate the content validity of the draft instrument and improvement.**

7. Conduct the tryout or the pilot study of the instrument for determining the reliability

Quality of Instrument

- ***Validity:*** the instrument really measures what we want to measure, i.e. it really measures the intended performance objective
- ***Reliability:*** the instrument provides the consistent results
- ***Accuracy:*** the instrument approximates an individual's true knowledge or ability
- ***Objectivity:*** the instrument provides the same meaning and understanding regarding the test question and the test score
- ***Practicality:*** the instrument is relevant to the practical situation
- ***Efficiency:*** the instrument can provide the maximum information with the minimum resource and time

Determining the Validity of the Instrument

- **The types of validity** *are* Content Validity, Construct Validity, Concurrent Validity, and Predictive Validity.
- Validity has to do with the meaning of the scores and the way we use the scores to make decision – whether the scores represent the characteristics we intended to measure.
- Example of invalid test scores -- the test is intended to measure higher-order thinking skills in science, but most items require only recall of facts, terms, and scientific principles – i.e. the meaning of scores has been distorted.
- Validity refers to the inferences drawn, not the instrument.
- Validity is inferred from available evidence (not measured)
- Validity depends on many different types of evidence

  Determine the validity of assessment results in terms of four types of evidence related to such assessment results

  1.Content-Related Evidence: How adequately does the sample of assessment tasks represent the domain of tasks to be measured?

2.Criterion-Related Evidence: How accurately does performance on the assessment (e.g., test) predict future performance (predictive study) or estimate present performance (concurrent study) on some other valued measure called a criterion?

3.Construct-Related Evidence

How well can performance on the assessment be explained in terms of psychological characteristics?

4.Consequences of Using Assessment Results

How well did use of the assessment serve the intended purpose (e.g., improve performance) and avoid adverse effects (e.g., poor study habit)?

How to Evaluate Content Validity

- The most important validity is the **content validity** --- *the extent to which the instrument measures what to be measured.* **The Expert Judgment** is the practical approach to estimating the content validity. At least **five experts** should be asked to judge each item whether it really measure the expected attribute.

- **The Item Objective Congruence (IOC) Index** is used as the basis for screening the item quality.

In each item, the experts are asked to determine the content validity score:

The score = 1, if the expert is sure that this item really measured the attribute.

The score = –1, if the expert is sure that this item does not measure the attribute.

The score = 0, if the expert is not sure that the item does measure or does not measure the expected attribute.

The Example of IOC Form (Four items were judged by three experts)

Item No.	Expert 1	Expert 2	Expert 3	Total Score	The IOC Index Mean of Expert Scores
1	1	1	0	2	IOC1 = 2/3 = .67
2	0	1	-1	0	IOC2 = 0/3 = 0
3	-1	1	1	1	IOC3 = 1/3 = .33
4	-1	0	-1	-2	IOC4 = -2/3 = -.67

The qualified items should have the IOC equal to or greater than 0.50

The Reliability of the Instrument

- Reliability (r) refers to the consistency of an assessment's results under different conditions. The reliability coefficients vary from 0.00 to 1.00, the desirable values should be .80 or above

- Four methods are used to determine reliability of test scores:

1.Test-retest method: Reliability is the stability of test scores over a given period of time. (r = correlation between test-retest scores)

2.Equivalent-forms method: Reliability is the consistency of the test scores over different forms of the test.

3.Test-retest with equivalent forms: Reliability is the consistency of test scores over both a time interval and different forms of the test.

4.Internal-consistency methods: Reliability is the consistency of test scores over different parts of the test. The well-known procedures of reliability estimation are:

4.1 Split-half method – correlation of sores obtaining from two parts of a test

4.2 Kuder-Richardson Formula 20 (KR-20)

4.3 Cronbach's Alpha Coefficient

Mostly, the Cronbach Alpha Coefficient (α) is applied for determining the reliability of the instrument.

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\Sigma (S_I)^2}{(S_T)^2}\right]$$

Where, n is the number of items
$(S_I)^2$ is the variance of each item
$(S_T)^2$ is the variance of total score

Traditionally, the criteria of reliability interpretation are as follows:

Reliability coefficients which are equal to or above 0.80, are considered as high reliability

Reliability coefficients which are between 0.60 to 0.79, are considered as moderate reliability

Reliability coefficients which are below 0.60, will be considered as low or poor reliability.

Relationship between Validity and Reliability

- It is possible for scores obtained from an instrument to be reliable (consistent) but not valid (measuring something else).
- In contrast, scores cannot be both valid and unreliable --- if scores measure what was intended for, it is implied that they will do consistently.
- A valid test is always reliable, but a reliable test is not necessarily valid.

Mode of Assessment

1.Observation Mode
It is the structured and systematic observation accompanied with the Data Record Forms like a checklist, rating scale, and description.

2.Self-Report, Self-Evaluation or Individual Opinion Mode
The examples of the instrument are Structured or Informal Interview with the Data Record Form, Questionnaire, Opionionnaire, Attitude Inventory, Interest Survey.

3.**Report by the Others** who know the subjects well. The information may be obtained through the interview or the use of a questionnaire with the people who know the student. The report about a student's behavior may come from the peer, teachers, and the parents.

Typical Instrument

1. Likert Scale

The Likert Scale is designed to capture a person's level of agreement with statements that present a position. For example:

Endorsement scale 1:

- *Strongly Agree*
- *Agree*
- *Uncertain*
- *Disagree*
- *Strongly Disagree*

Endorsement scale *2:*

- *Definitely True*
- *True*
- *Don't Know*
- *False*
- *Definitely False*

Frequency scale:

- *Always*
- *Very Often*
- *Sometimes*
- *Rarely*
- *Never*

2. Semantic Differential Scale.

- Measuring attitudes towards certain targets.
- Each target will be followed by a set of paired adjectives.

The example of paired adjectives:

Good-Bad	*Beautiful-Ugly*	*Clean-Dirty*	*High-Low*	*Valuable-Worthless*
Kind-Cruel	*Pleasant-Unpleasant*	*Bitter-Sweet*	*Happy-Sad*	*Relaxed-Tense*
Clear-Hazy	*Rough-Smooth*	*Rich-Poor*	*Healthy-Sick*	*Bright-Dark*
Nice-Awful	*Fair-Unfair*	*Empty-Full*	*Tasty-Distasteful*	

- Put one target a time.
- Certain number of paired adjectives placed at the end of continuum.
- The scale can be five-, seven-, nine-, or eleven-scale score.

For example,

MATHEMATICS

Bad 1----------2----------3----------4----------5----------6---------7 Good
Unfair1----------2----------3----------4----------5----------6---------7 Fair
Dirty 1----------2----------3----------4---------5----------6----------7 Clean
Sad 1----------2----------3----------4----------5----------6---------7 Happy

ADMINISTRATOR

Inattentive	1……. 2……. 3……. 4……. 5	Attentive
Worthless	1……. 2……. 3……. 4……. 5	Valuable
Unfair	1……. 2……. 3……. 4……. 5	Fair
Dirty	1……. 2……. 3……. 4……. 5	Clean

Limitation of Non-test Instrument, especially the self-report instrument

1. Definition problems lead to unacceptable measurement.
2. Observations could be confound with bias, e.g. Halo Effect, Hawthorne Effect.
3. Inventory scales -- self-report -- tend to be interfered by faking, response set, acquiescence, social desirability, etc.
4. Assessment of affective characteristics may raise ethical issues, e.g. invasion of privacy.

Strategy for controlling the errors inn data from self-report measures:

1. *Assurance of privacy and confidentiality* – communicating to respondents that results will not be shared with anyone other than a select group of individuals whom the respondents trust; making sure that the confidentiality is maintained. Respondents feel better if the data will be used for development; no negative effect to personal well-being; the individual data will be not disclosed to anyone.
2. *Obtaining informed consent* – openly communicating to respondents the purposes for assessment and how the data will be used; asking for the cooperation and informed consent of respondents in gathering accurate data to fulfill the purposes.
3. *Randomly mixing of negatively and positively stated items.*
4. *Applying triangular measurement*

Scoring the Negative Items

Usually, items of both directions are used to provide more variety and breadth in the items, and crossed check the true characteristics. For example,

Positive item: “Mathematics is useful for future work”

Negative item: “It is boring to study mathematics”

In the five-point scale, the score of negative item is (6 – response point).

Scale Score of Positive Item

Rating	Rating Score	Scale Score
Strongly Agree	5	5
Agree	4	4
Undecided	3	3
Disagree	2	2
Strongly Disagree	1	1

Scale Score of Negative Item

Rating	Rating Score	Scale Score
Strongly Agree	5	1
Agree	4	2
Undecided	3	3
Disagree	2	4
Strongly Disagree	1	5

Interpretation of Scale Scores

Based on the five-point scale measuring the degree of the characteristics such as attitude, satisfaction, or perception, the mean scale score mostly varies from 1.00 to 5.00. The following criteria of scale interpretation will be used:

Scale Score	Meaning
4.50 - 5.00	Very High
3.50 – 4.49	High
2.50 – 3.49	Moderate
1.50 – 2.49	Low
1.00 – 1.49	Very Low

Example of Rating Scale

- Mathematics is interesting subject.
    a. Strongly Agree.
    b. Agree.
    c. Neutral.
    d. Disagree
    e. Strongly Disagree.
- Do you like to mix with people socially?
  5.Almost always
  4.Frequently
  3.Occasionally
  2.Rarely
  1.Almost never
- After the learning session, do you go to the library for further study?
  1.Usually.
  2.Frequently.
  3.Occasionaly.
  4.Seldom.
  5.Rarely.

- How is the creative ability of this student?
  1.Outstanding.
  2.Above average.
  3.Average.
  4.Below average.
  5.Unsatisfactory.
- How active is the student in class project?

  Least 1 2 3 4 5 Most
- How well does the student relate to his/her peer*?*

  Least 1 2 3 4 5 Most
- How enthusiastic was your instructor in presenting the course materials?

  Very Unenthusiastic **1 2 3 4 5** *Very enthusiastic*
- Generally, how attentive were your teacher in class?

  Very inattentive **1 2 3 4 5** *Very attentive*
- I enjoy solving complicated problem.

  a. Strongly agree or absolutely true.
  b. Rather agree or mostly true.
  c. Undecided.
  d. Rather disagree or mostly untrue.
  e. Strongly disagree or absolutely untrue.
- How high would you rate the quality of math teaching?
  a. Very high
  b. High
  c. Average
  d. Low
  e. Very low
- Rate the quality of school services.
  a. Very good
  b. Good
  c. Satisfactory
  d. Poor
  e. Very poor
- To what extent did you learn from biology course?
  5. To a great extent
  4. To a moderate extent
  3. To some extent
  2. Very little
  1. Not at all

- Describe how much you like cultural program?
    a. Like it a lot
    b. Like it somewhat
    c. Not sure
    d. Dislike it somewhat
    e. Dislike it a lot
- How often do you submit the homework on time?
    a. All the time
    b. Most of the time
    c. Some of the time
    d. Rarely
    e. Not at all

Training Need Survey for Teachers

Directions: For the item listed, you will be asked to indicate the level of adequacy of your previous training in the different areas of teaching. Your response choices for all the items are as follows:

a. More than adequate
b. Adequate
c. Less than adequate
d. No training at all

ITEMS	**RESPONSES**
1.Planning for instruction using knowledge of what students should learn.	a-b-c-d
2.Planning for instruction using knowledge of subject matter.	a-b-c-d
3.Planning for instruction using instructional design principles.	a-b-c-d
4.Developing the curriculum using community values.	A-b-c-d
5.Using classroom assessment to plan and improve teaching.	a-b-c-d
6.Maintaining high standards of classroom conduct/behavior.	a-b-c-d
7.Designing activities that support academic growth	a-b-c-d
8.Designing activities that develop students' interpersonal skills	a-b-c-d
9.Designing activities that build students' critical thinking skills	a-b-c-d
10.Developing activities that use modern media/technology	a-b-c-d

Example of Observation Scale

Class: Grade 10 Chemistry Laboratory.

Measurement Objectives: Responsibility Behaviors.

Instruction: Each student is observed weekly within 2-hour use of laboratory for 4 weeks. The individual records can be transferred to the whole class record. The following scale is used for each behavior judgment.

1. *Not at all*
2 .Rarely
3. Sometimes
4. Most of the time
5. All the time

Observation Form

Student's Responsibility Behavior

Student Name ----------

Item	1	2	3	4	5
1. Attending the class on time. 2. Bring all required materials. 3. Participate the group planning. 4. Perform the tasks as designed by the leader. 5. Share and exchange ideas when needs arisen. 6. Help others in the group for common achievement. 7.Take good care of apparatuses without damage or waste. 8. Clean the apparatuses, table, and used area.					