Chapter 2.

What to Measure and How to Interpret Results

Contents

- What to measure

- Two Influential Interpretive Alternatives
  - Norm-referenced Interpretation and
  - Criterion-Referenced Interpretation
  - Three kinds of Outcomes(Test Domain)

- Three kinds of Decisions

What to Measure

- Students are supposed to learn certain things

- A teacher is supposed to assess students to see if what they were supposed to learns has been learned.

What to Measure

- 1) Textbooks
- 2) Content Standards
- 3) Teachers' Preferences

=> Three major sources for

measurement

=> Things you have to concern when making test items.

1) Textbooks

- You cannot measure all the things in a textbook!

- Time for

Assessing + Teaching= 1(fixed)

2) Content Standards

- Refer to the <u>knowledge and skills</u> educators want students to learn.

  e.g.) ability to communicate effectively in writing.

- Objective = Content Standard

2) Content Standards

- Performance Standards

: refer to the <u>level of proficiency</u> at which students are supposed to master content standards.

2) Content Standards

- What's different between
Content Standards
and Performance Standards?

?

2) Content Standards

- **EXAMPLE: A Content Standard**

  Mathematics Standard 11: Statistics and Probability

  …in grades K-4(4 years old), the mathematics curriculum should include experiences with data analysis and probability so that students can--

- ◆ collect, organize, and describe data;

- ◆ construct, read, and interpret displays of data;

  formulate and solve problems that involve collecting and analyzing data; and

- ◆ explore concepts of chance.

  Source: National Council of Teachers of Mathematics, *Curriculum and Evaluation Standards for School Mathematics*. (www.Ed.gov U.S. Department of Education)

2) Content Standards

- **EXAMPLE: Performance Standards**

Reading : *Reading is a process which includes demonstrating comprehension and showing evidence of a warranted and responsible interpretation of the text. "Comprehension" means getting the gist of a text. It is most frequently illustrated by demonstrating an understanding of the text as a whole; identifying complexities presented in the structure of the text; and extracting salient information from the text. In providing evidence of a responsible interpretation, students may make connections between parts of a text, among several texts, and between texts and other experiences; make extensions and applications of a text; and examine texts critically and evaluatively. * * **

2) Content Standards

E1d The student reads aloud, accurately (in the range of 85-90%), familiar material of the quality and complexity illustrated in the sample reading list, and in a way that makes meaning clear to listeners by--

◆ self-correcting when subsequent reading indicates an earlier miscue;

◆ using a range of cueing systems, e.g., phonics and context clues, to determine pronunciation and meanings; and

◆ reading with a rhythm, flow, and meter that sounds like everyday speech.

Some examples of activities through which students might produce evidence of reading aloud accurately:

◆ Reading aloud to peers or younger children.

◆ Participating in a Readers' Theater production.

◆ Recording an audiotape or videotape an example of reading aloud.

Source: *New Standards, Performance Standards, Volume 1 - Elementary School,* (National Center on Education and the Economy and the University of Pittsburgh, 1997) (www.Ed.gov U.S. Department of Education)

3) Teachers' Preferences

⊙ Insufficient time left for instructional purposes if a teacher wants to assess everything taught.

=>> Deciding which contents will be included is needed.

Two Influential Interpretive Alternatives

1) Relative, Norm-Referenced Interpretations

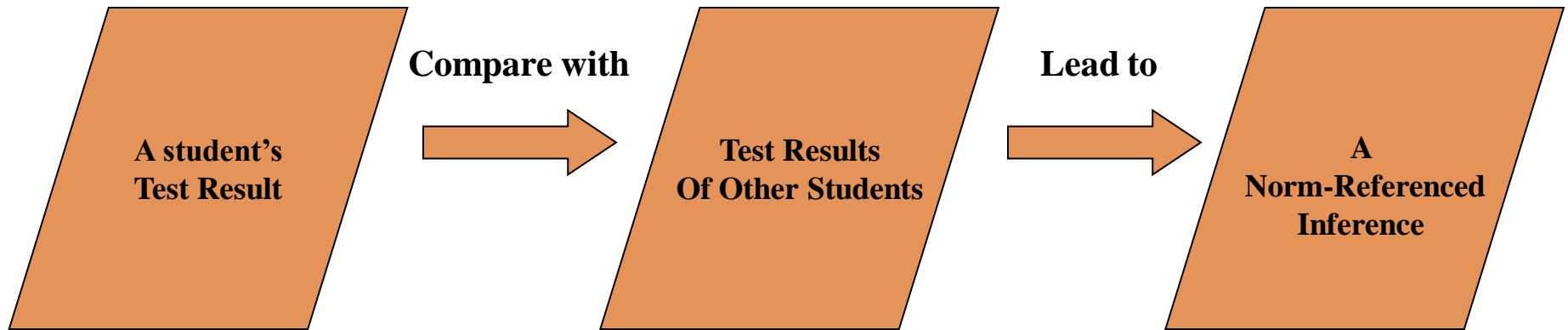2) Absolute, Criterion-Referenced Interpretations

Before we start.....

- The chief determiner of *what* should be measured rests on *why* you're measuring anything in the first place.

- Test Scores in and of themselves
  do not make any sense at all!!
  (Good example on P30, second paragraph)

- What educators need are results-based interpretations that have implications for educational decision

1) Relative, Norm-Referenced Interpretations

- Comparing one student's test results with that of the student in the norm group.

  => Purpose is to determine 'Who is better?'

- Raw score < %(Percentile)

- A test is not norm-referenced, but the results-based interpretation you make is norm-referenced.

- Entrance Exam
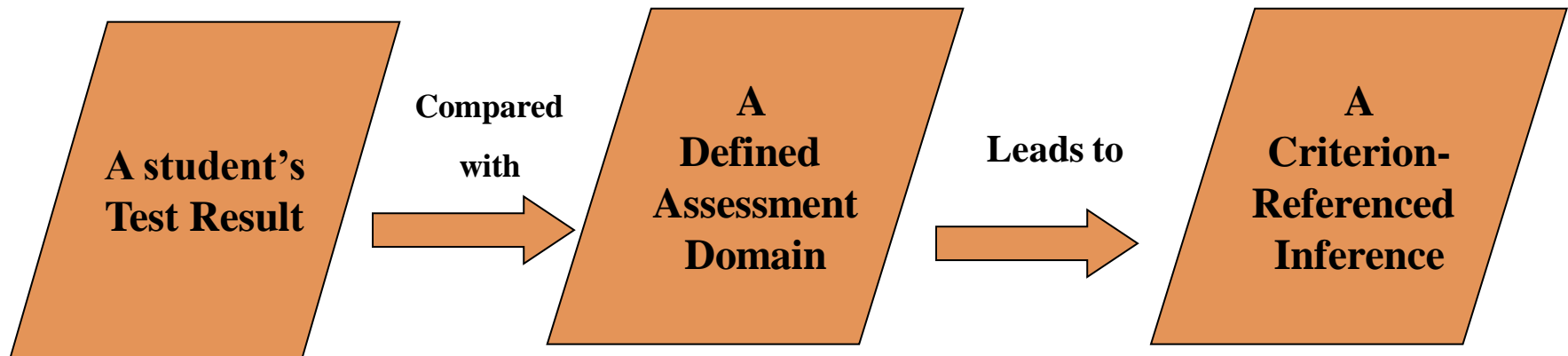
1) Relative, Norm-Referenced Interpretations

A student's Test Result → **Compare with** → Test Results Of Other Students → **Lead to** → A Norm-Referenced Inference

A relative interpretation of a student's assessment performance.

2) Absolute, Criterion-Referenced Interpretations

- Giving meaning to a test result by comparing it with a defined assessment domain.

=> Purpose is to determine what students can or can't do.

- Raw score> %(Percentile)

- Clearly described assessment domain is needed.

- Graduation exam

2) Absolute, Criterion-Referenced Interpretations

A student's Test Result

Compared with

A Defined Assessment Domain

Leads to

A Criterion-Referenced Inference

An absolute interpretation of a student's assessment performance.

3) Different Ancestries

* Which one appeared first ?

   Norm-Referenced vs Criterion-Referenced

** **Tip**

   : Try to recall the Chinese civil service exam.

3) Different Ancestries

- 70s Robert Glaser: introduced the concepts of norm-referenced and criterion-referenced measurement(1963)

=> Prevalence of criterion-referenced measurement

- Both norm-referenced and criterion-referenced interpretations are needed if educators are to accomplish the full range of necessary purposes.

Three Kinds of Outcomes

- Cognitive test: knowledge or intellectual skills

    (1) achievement tests: current knowledge
    (2) aptitude tests: potential to perform well subsequently.

- Psychomotor test: physical competencies

- Affective test: Attitudes, interests, and values

Three Kinds of Outcomes

- Cognitive and Psychomotor Assessment
  : optimal performance

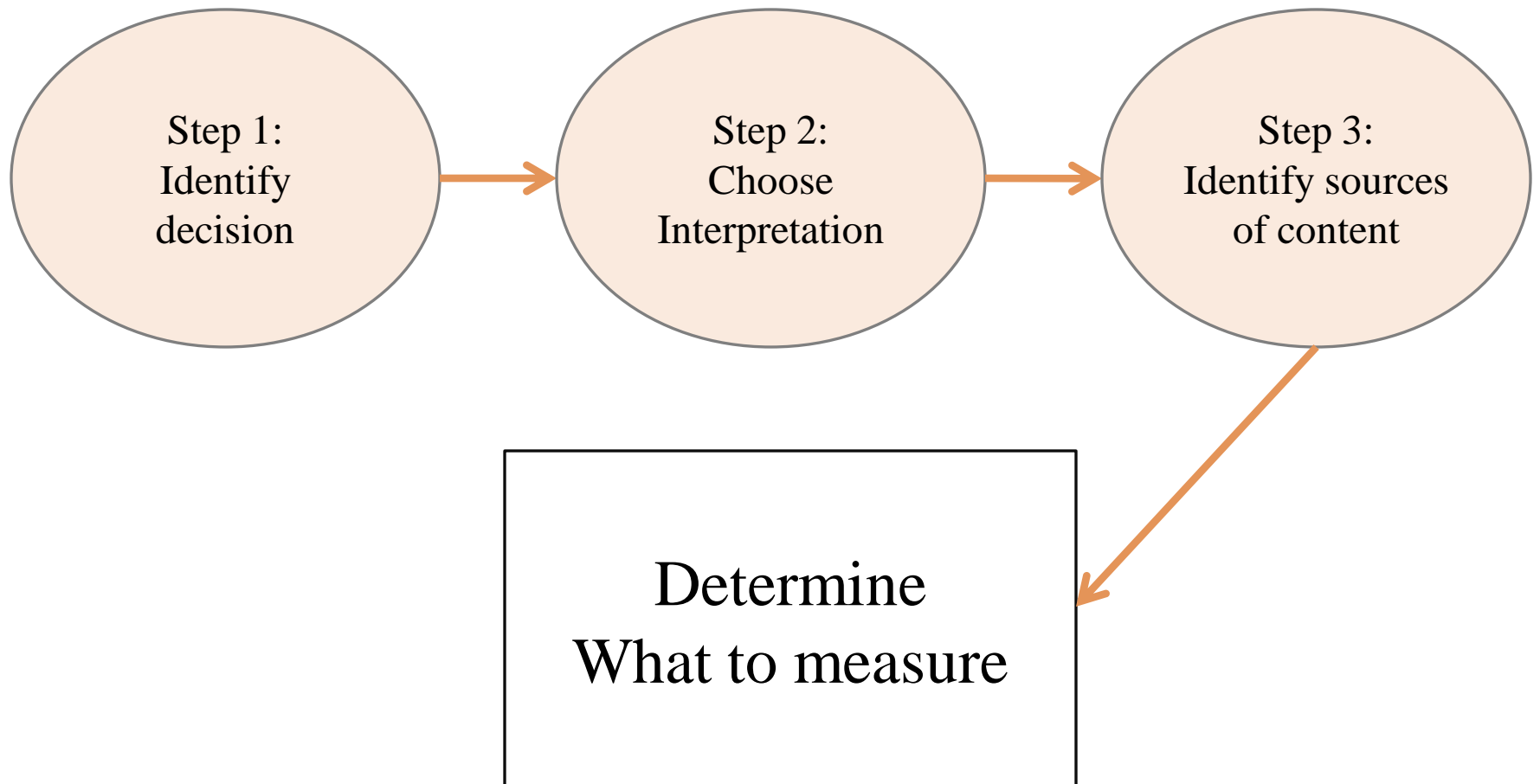 " How well can you perform this skill?"

- Affective Assessment
  : typical performance

  " What will you do?"

→ For affective measure, the distinction between aptitude and achievement makes little sense.

Three kinds of Decisions

A three–step approach to the determination of what should be measured

Step 1:
Identify
decision

Step 2:
Choose
Interpretation

Step 3:
Identify sources
of content

Determine
What to measure

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

1. Fixed-quota settings

- ✓ Quota : the limited number of those who are officially allowed

- ✓ Situations where there are more applicants than openings

- ✓ For example, prestigious law school 100 openings, over 500 applicants

1. Fixed-quota settings

✓ In the fixed-quota setting, it is necessary to sort out individuals according to their relative abilities.

✓ Norm-referenced interpretations are appropriate.

✓ Who is best or worst

2. Requisite-skill/knowledge settings

- ✓ Situations where the focus of the decision is who is qualified

- ✓ For example, medical school

  - ▪ "everything-a-patient-would-like-you-to-know" examination

  - ▪ Only those candidates who display the requisite medical knowledge and skills should be certified to practice.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

2. Requisite-skill/knowledge settings

- ✓ In this setting, norm-referenced inferences would not be appropriate.

- ✓ Criterion-referenced inferences should be employed.

- ✓ Another example, the licensing of educators
  - ▪ Truly qualified teachers should be employed

3. Dominant score interpretations

✓ The majority of tests designed to make a selection decision are used in fixed-quota contexts.

✓ For selection decision, the type of inference depends on whether a fixed-quota or requisite-skill/knowledge setting is involved.

4. Content sources

✓ Content refers to the major ingredients contained in ameasuring device.

✓ Examples

- reading achievement test

- Biology test

- Psychomotor test

- Affective assessment device

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

1. Evaluation of instruction

✓ When an educator carry out instructional evaluation, there are three decision options.

3. Discard it

If the instruction is determined to be downright dismal

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

2. Dominant score interpretations

✓ For evaluating instruction, the most useful kind of interpretation is a criterion-referenced interpretation because educators are interested in the degree to which students have learned particular knowledge and skills.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

2. Dominant score interpretations

✓ For the evaluation of students, it is difficult to say which kind of assessment-based inference will be appropriate.

✓ In short, the grading rationale should be identical to the type of results-based interpretations that the teacher uses.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

3. Content sources

✓ If instructional evaluation is taking place, the content of the assessments needs to be reflect, "What *should* be taught to students at this age?"

✓ For evaluating instruction, the content for a test must be chosen by subject-knowledgeable teachers and curriculum specialists.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

3. Content sources

✓ For the evaluation of students, the appropriate question to ask is "What *has* been taught to these students by this teacher?"

✓ The content of grade-determining tests should reflect the content covered during the teacher's instruction.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

1. What to teach?

- -

✓ A skilled teacher will make instructional decisions by using

evidence from pre-assessments that identify the skills and knowledge

with which the students enter the class.

2. How long to teach?

✓ The teacher's decision to stop instruction aimed at a particular objective might be made on a student-by-student basis.

✓ Progress-monitoring assessments help teachers decide how long to give instruction.

Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

3. Dominant score interpretations

✓ For instructional decisions, absolute, criterion-referenced interpretations of students' assessment performances because they tell what it is that students can or can't do.

✓ Relative inferences are less useful because the norm group often shifts or is unknown.

4. Content sources

- ✓ In deciding what to teach, teachers will rely on their own subject-matter expertise and engage in the determination of curriculum.

- ✓ Curriculum : The ends, that is, the learning objectives sought for students

- ✓ The objectives should be age-appropriate instructional.
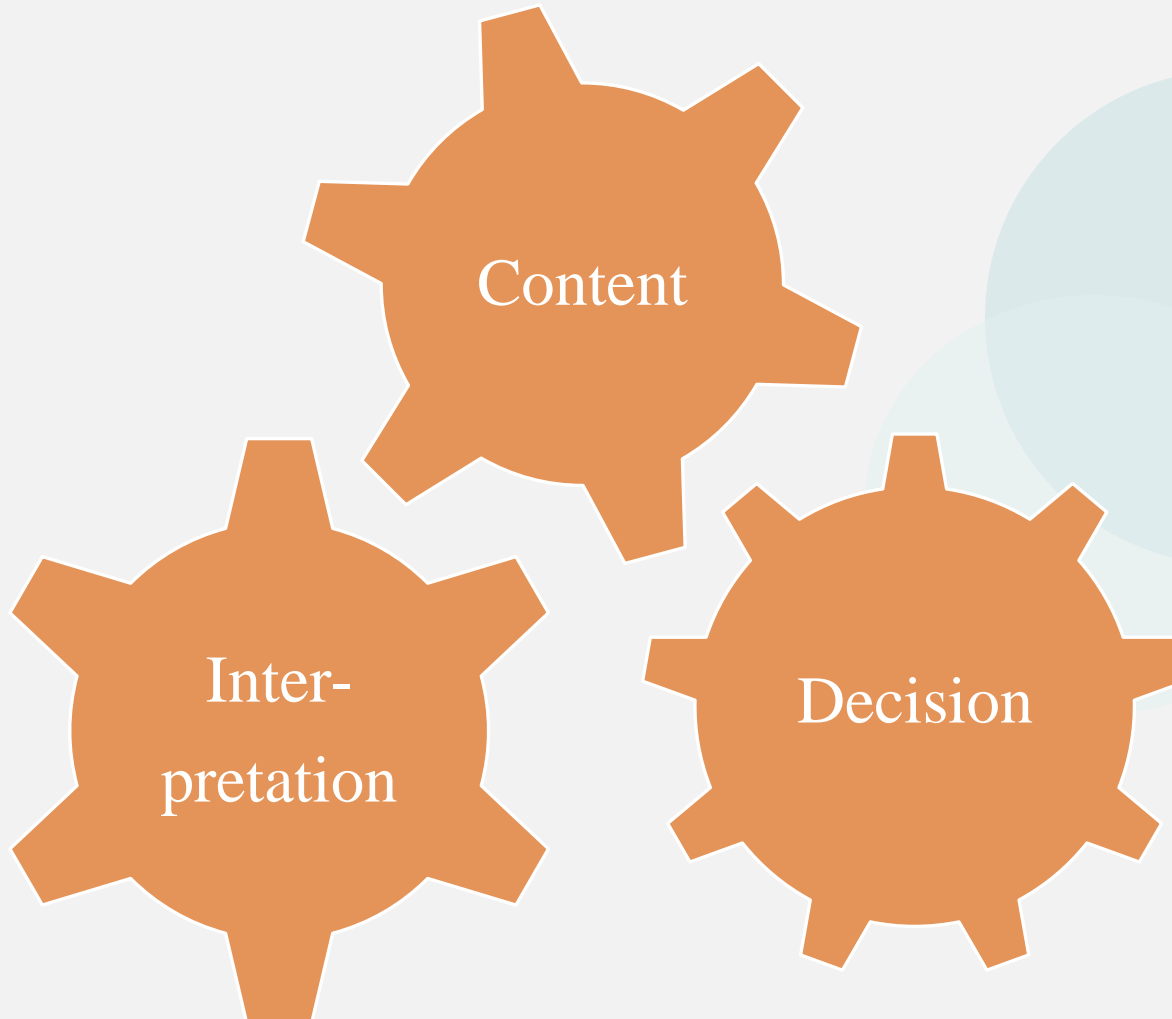
Selection
Decisions

Evaluation
Decisions

Instruction
Decisions

4. Content sources

✓ In deciding how long to teach, progress-monitoring assessments can be employed and devised to see when to stop instructing.

✓ Any project-monitoring assessments must be based on the current objectives teachers pursue.

Factors to consider in deciding what to measure

Content

Inter-pretation

Decision

Thank you.