SPAU 328

Principles of Evaluation, Diagnosis, and Report Writing in ComD

Dina Budeiri MSc

Types of assessment

Two general types of assessments

Standardized assessment

Non-standardized assessment

Standardized Assessment

- Systematic
  - *Say* and *Do* controlled (by manual)
  - Specific stimuli are used
  - Explicit rules for scoring the test
  - Not influenced by the examiner's biases
  - Measurement process will be uniform across examiners

Non-Standardized Assessment

- Questionnaires: administered personally or sent to clients, parents, teachers,..

- Questionnaires may be more valid than formal tests! Interviews (stand alone or in conjunction with questionnaires).

- Interviews are more personalized than questionnaires

- The use of questionnaires & interviews —> quantitative data and qualitative data (verbal descriptions)
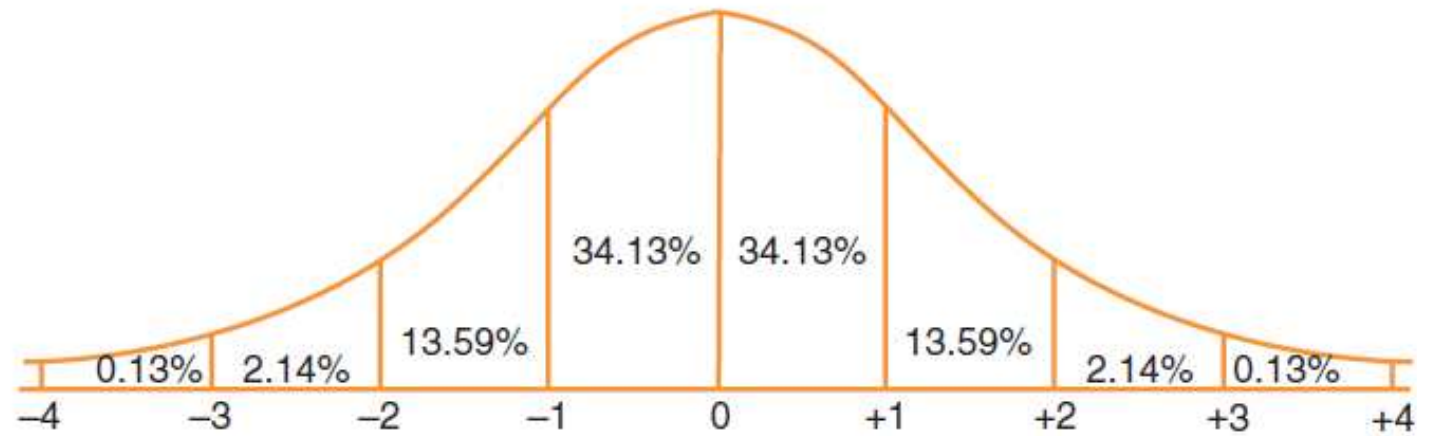
Assessment methods

- Norm-referenced tests
- Criterion-Referenced Tests
- Authentic assessment
  - Dynamic assessment
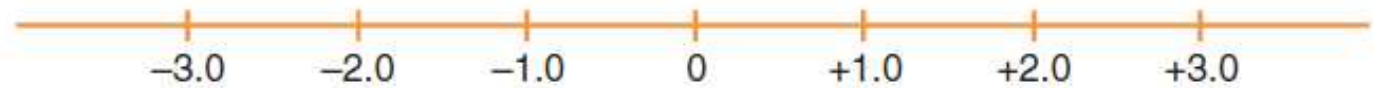
Norm-referenced tests

- Norm-referenced tests are always standardized. They allow a comparison of an individual's performance to the performance of a larger group, called a **normative group.**

- The results of the normative group are used to create **normal distribution curve**

% of people who will score within each standard deviation

34.13%   34.13%

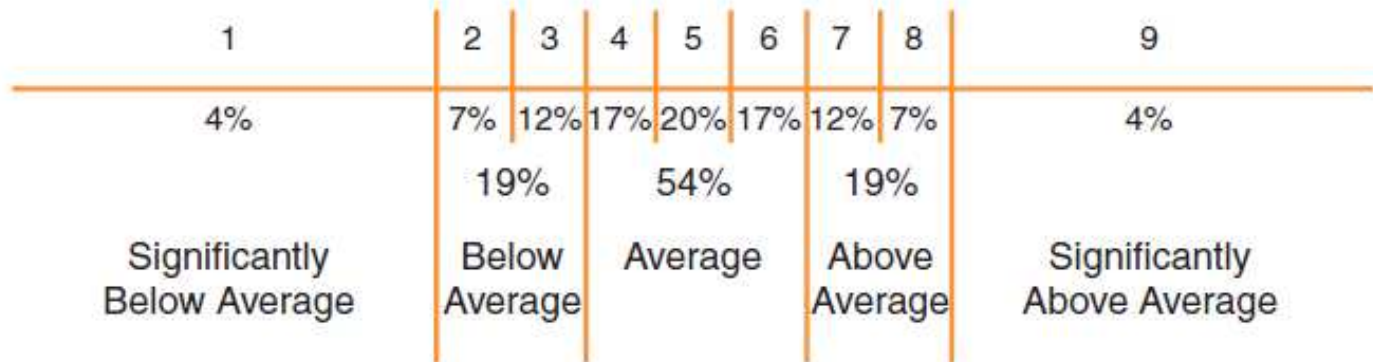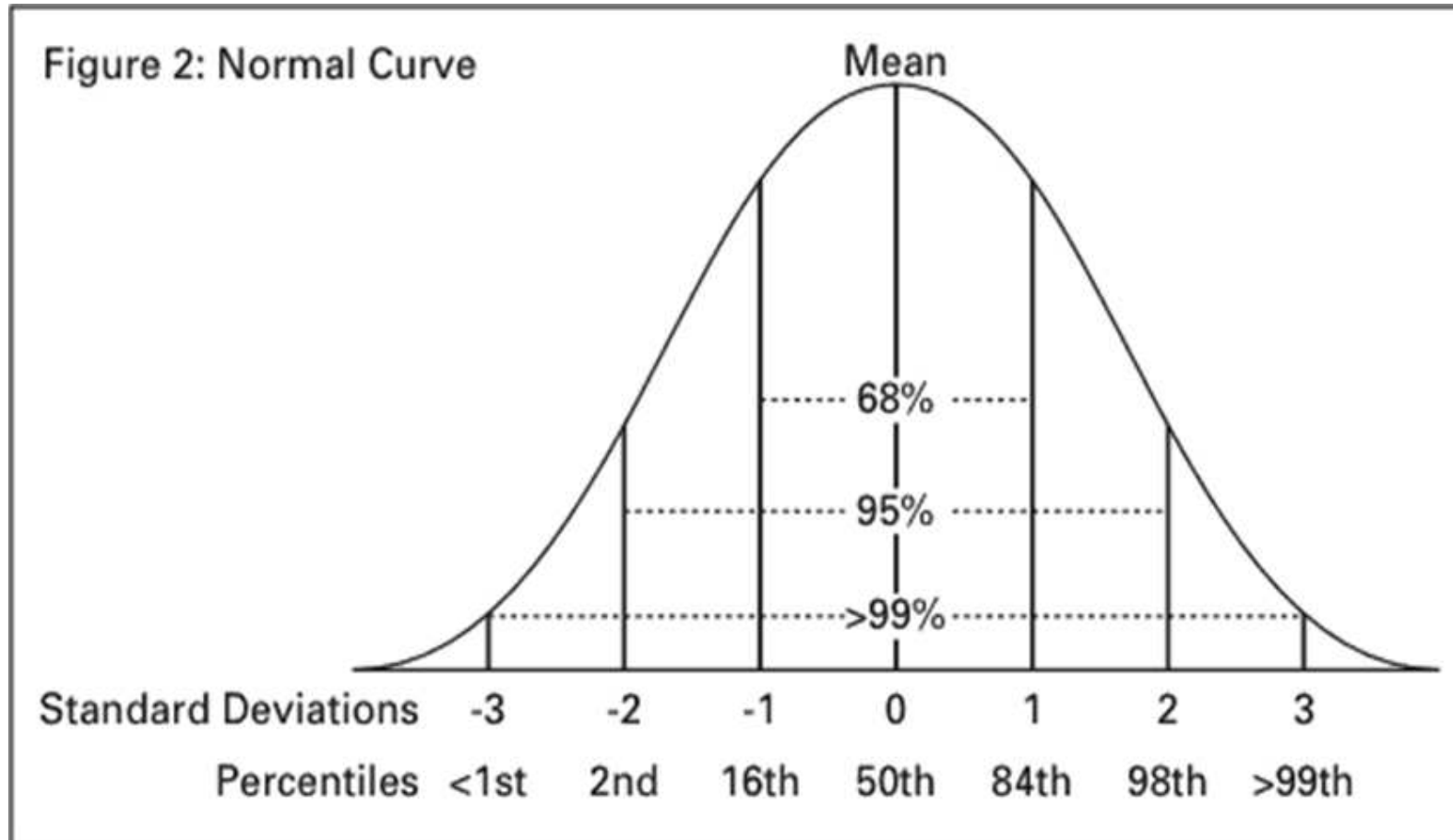13.59%   13.59%

0.13%   2.14%   2.14%   0.13%

Standard deviation

−4   −3   −2   −1   0   +1   +2   +3   +4

z-score

−3.0   −2.0   −1.0   0   +1.0   +2.0   +3.0

Percentile rank

1   5   10 20 30 40 50 60 70 80 90   93   99

Stanine

1   2 3 4 5 6 7 8   9

% of people who will obtain the corresponding stanine

4%   7% 12% 17% 20% 17% 12% 7%   4%

19%   54%   19%

Significantly Below Average   Below Average   Average   Above Average   Significantly Above Average

Figure 2: Normal Curve

Types of Scores in Standardized Assessment

Standardized testing —> scores (quantitative measures)

A test administration —> raw score

- **Raw score** is the actual scores earned on a test
- **Raw score** —> converted to be viewed on a **distribution**
- **Distribution** —> measures of the *client's performance* compared to the *performance of the norming sample*
- Two statistical measures of a **distribution**:
  - mean (the arithmetic average of the scores of the norming samples)
  - standard deviation (the extent to which scores deviate from the mean)

The client's score(s) can be expressed in two ways:

- **standard deviation** from the mean

- **percentile rank** (use percentile points to express a child's score relative to the norming sample)

i.e., if a child scores is at the 25th percentile, 75% of the children will have done better on the test than that child did
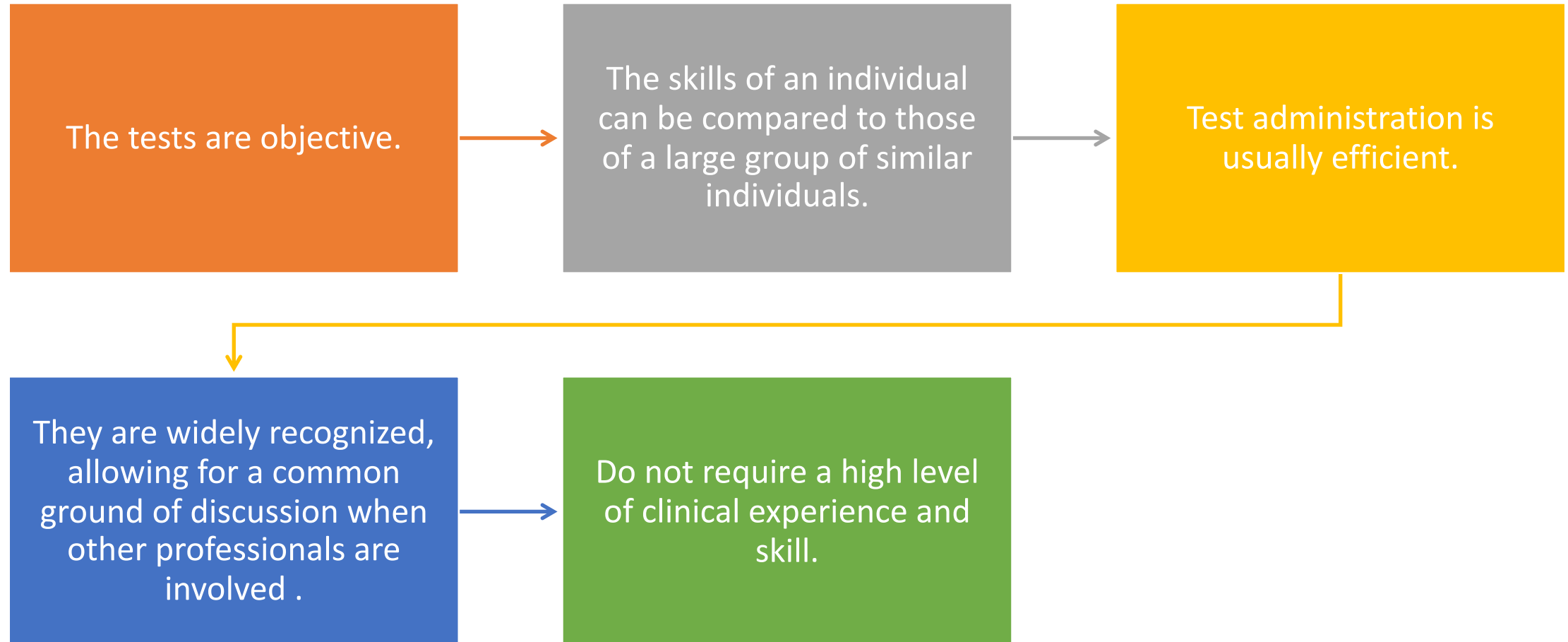
The **median** is the score in the exact middle of the distribution
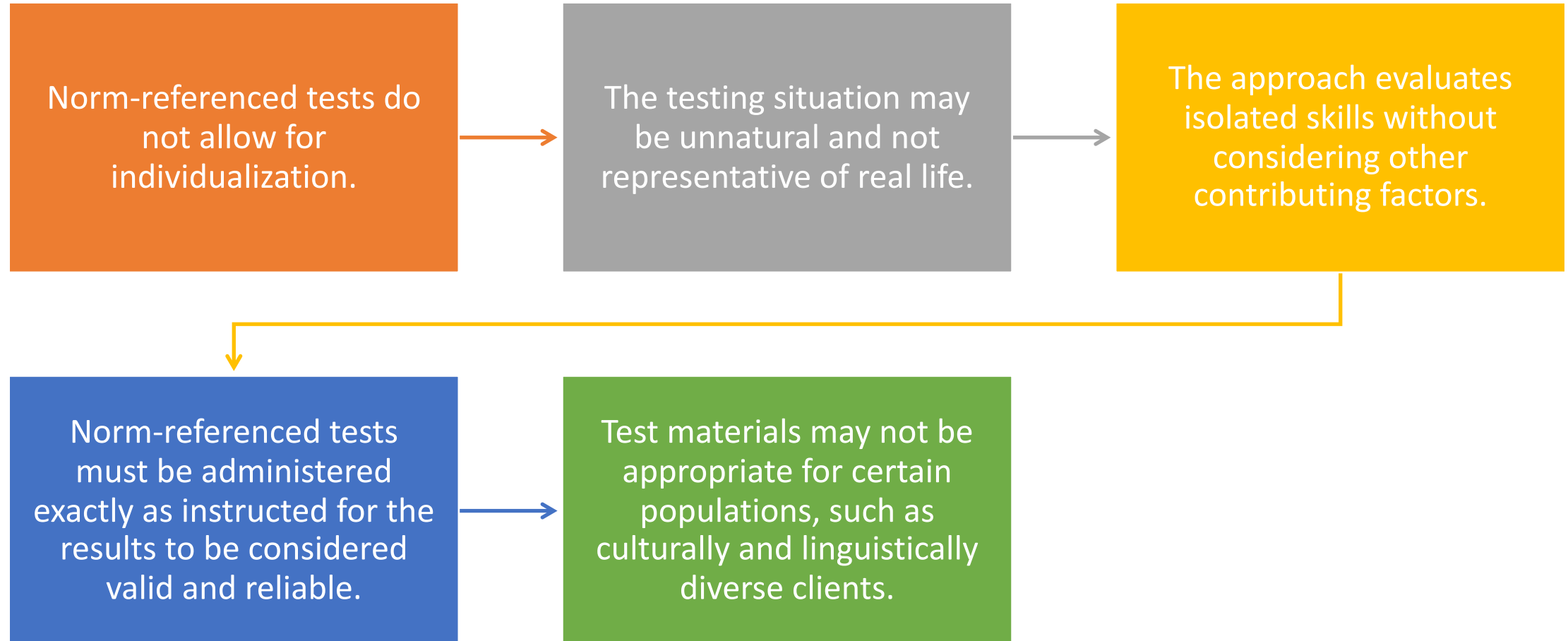
Measurement Scales:

nominal scale: a category is *present* or *absent* (e.g., hoarseness,...).

ordinal scale: a *numerical* scale that can be arranged according to rank orders or levels.

Norm-referenced tests | advantages

The tests are objective. → The skills of an individual can be compared to those of a large group of similar individuals. → Test administration is usually efficient.

They are widely recognized, allowing for a common ground of discussion when other professionals are involved . → Do not require a high level of clinical experience and skill.

Norm-referenced tests | disadvantages

Norm-referenced tests do not allow for individualization.

The testing situation may be unnatural and not representative of real life.

The approach evaluates isolated skills without considering other contributing factors.

Norm-referenced tests must be administered exactly as instructed for the results to be considered valid and reliable.

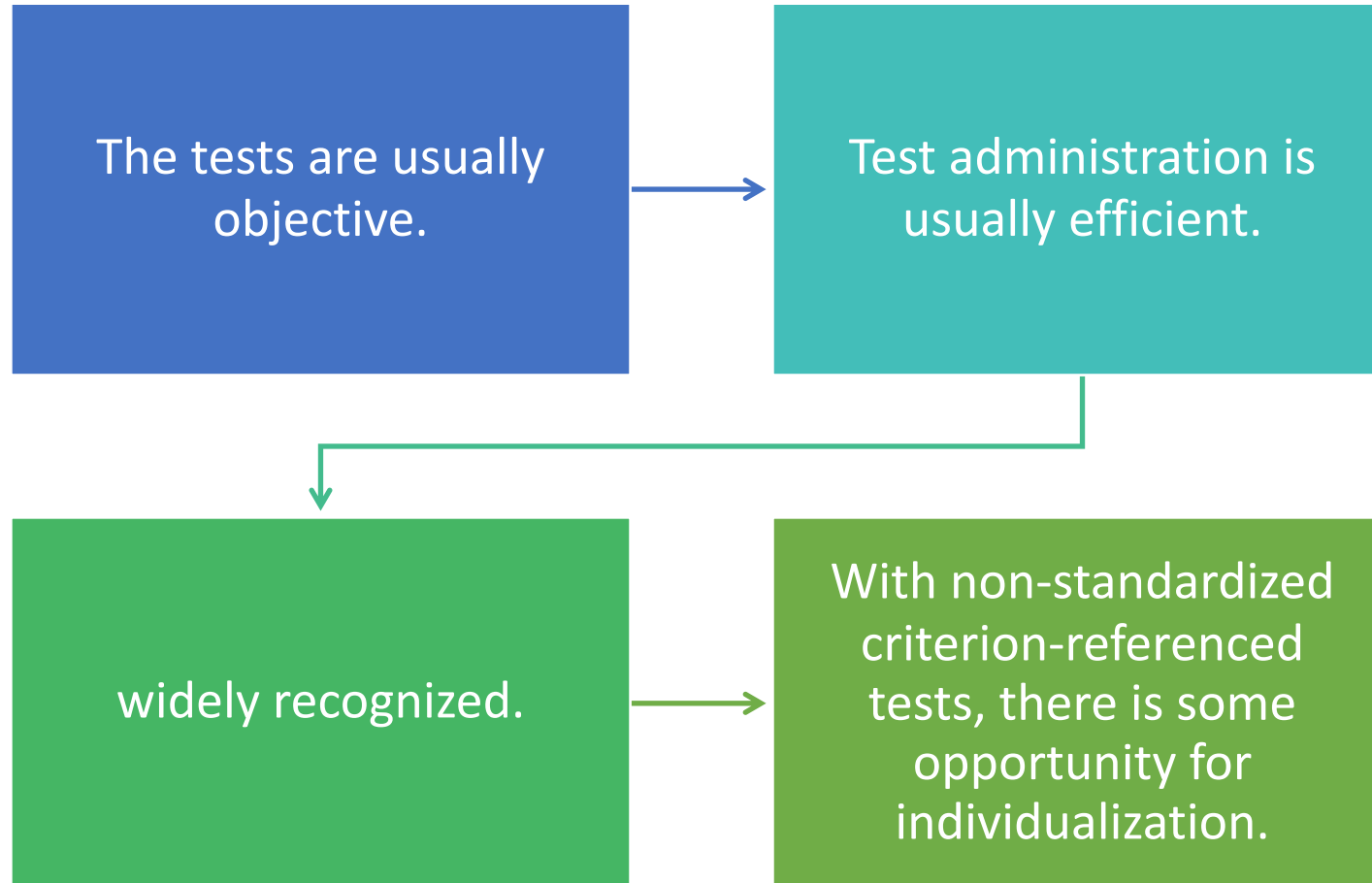Test materials may not be appropriate for certain populations, such as culturally and linguistically diverse clients.
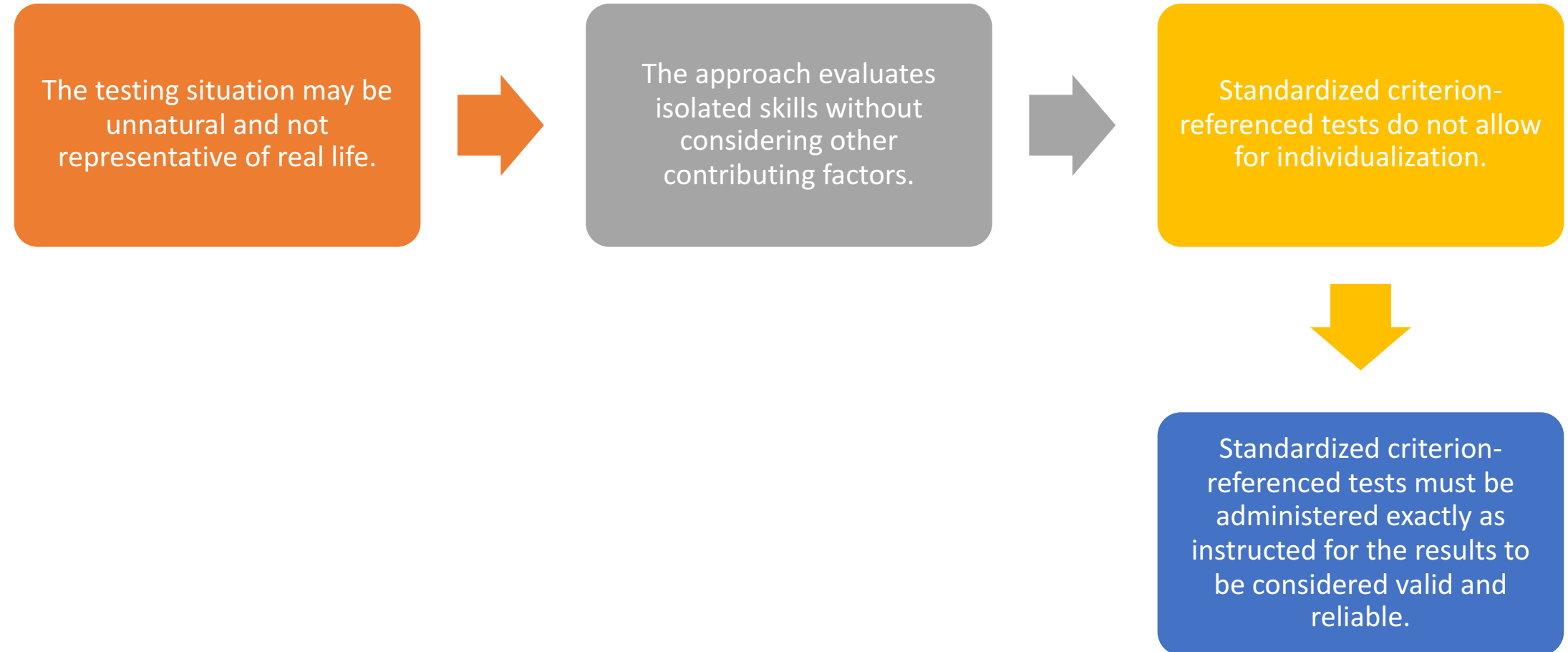
Criterion-Referenced Tests (CRT)

- CRTs identify what a client can and cannot do compared to a predefined criterion.

- CRTs assume that there is a level of performance that must be met for a behavior to be acceptable.

- Examples?

Criterion-Referenced Tests| advantages

The tests are usually objective.

Test administration is usually efficient.

widely recognized.

With non-standardized criterion-referenced tests, there is some opportunity for individualization.

Criterion-Referenced Tests| disadvantages

The testing situation may be unnatural and not representative of real life.

The approach evaluates isolated skills without considering other contributing factors.

Standardized criterion-referenced tests do not allow for individualization.

Standardized criterion-referenced tests must be administered exactly as instructed for the results to be considered valid and reliable.

Norm-referenced vs. criterion-referenced

	Norm-referenced	Criterion-referenced
Simple definition	Assessments report whether test takers performed better or worse than a **hypothetical average (Norm)** of a population, which is determined by comparing scores against the performance results of a statistically selected group of test takers, typically of the same age or grade level, who have already taken the exam. They are mostly used by SLPs	Assessments that are designed to measure one's performance against a fixed set of predetermined **criteria** or learning standards

	Norm-referenced	Criterion-referenced	
Standardized vs. non-standardized	Always standardized	Could be standardized or non-standardized	
Examples	❖ Mean length of utterance (MLU) ❖ Stuttering severity Instrument (SSI) ❖ Clinical evaluation of language fundamentals (CELF)	Standardized	Non-standardized
		Assessment, Evaluation, and Programming System for Infants and Children (AEPS)	Naturalistic observation

	Norm-referenced	Criterion-referenced
Advantages	❖ Allows clinician to determine if the clients has problem, if the problem is clinically significant, and if problem warrants intervention ❖ Results of the client can be compared to a large group of similar individuals	❖ Non-standardized C-RT can be individualized and therefore inclusive ❖ Pinpointing a student's present level of performance ❖ Helpful in writing goals and objectives ❖ Monitoring incremental progress through treatment
Disadvantages	❖ Assessment is not culturally and linguistically diverse to represent all segments of the society	❖ Standardized C-RT do not allow for individualization ❖ Evaluates isolated skills without considering other contributing factors

Authentic assessments
(Alternative/ Non-traditional)

- A form of assessment in which patients are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills

- Known as alternative or non-traditional assessment

- Identifies what a client can and cannot do

- The test environment is more realistic and natural

- Ongoing assessment process

- Assessment information is maintained in a client portfolio (file)

- Requires more skilled, experienced, and creative SLPs

- Involves caregivers and other professionals besides the client

- Non-standardized

Authentic assessment Cont.

- authentic assessment identifies what a **client can and cannot do** with an emphasis on

- **contextualized test stimuli**.

- The test environment is more realistic and natural.

- **it is ongoing.** The authentic assessment approach evaluates the client's performance during diagnostic and treatment phases.

- requires more clinical skill.

How do we do it?

Authentic assessment Cont.

- Systematic observations
- Real-life simulations
- Language sampling
- Structured symbolic play
- Short-answer and extended-answer responses
- Self-monitoring and self-assessment
- Use of anecdotal notes and checklists
- Videotaping
- Audiotaping
- Involvement of caregivers and other professionals

Authentic assessments| advantages

The approach **is natural and similar** to the real world

Clients participate in **self-evaluation** and self-monitoring

The approach allows for **individualization.** This is particularly beneficial with culturally diverse clients or special needs clients, such as those who use(AAC) systems

The approach offers **flexibility**

Authentic assessments| disadvantages

The approach may lack objectivity

Procedures are not usually standardized, thus reliability and validity are less assured

Requires a high level of clinical experience

The approach is not efficient, requiring a lot of planning time

Dynamic assessment

- Measures the client's ability to *learn over time* when provided with instruction

- Appropriate for culturally and linguistically diverse populations

- **A test-teach-retest method :**
    1. A test is administered without prompts or cues to determine current(baseline) performance
    2. The clinician teaches (Mediated Learning Experience MLE).
    3. The clinician re-administers the test

- The purpose of dynamic assessment is to evaluate a client's learning potential based on his or her ability to modify responses after the clinician provides teaching or other assistance

- It is highly interactive

Dynamic assessment

- Dynamic assessment allows the clinician, as part of the diagnostic process, to determine baseline ability and identify appropriate goals and strategies for intervention. If one of the clinician's purposes is to discern a language difference versus a language impairment, it is helpful to note that clients who do not demonstrate improvement following teaching likely have a language impairment, whereas clients who are able to make positive changes following brief teaching experiences are likely to have a language difference.

Psychometric principles

- Psychometrics refers to the measurement of human traits, abilities, and certain processes.

- It is what speech-language pathologists do when evaluating a client's communication.

Psychometric principles

Validity

Reliability

Standardization

Validity

- The degree to which a test measures what it purports to measure.
- i.e. (**P**eabody **P**icture **V**ocabulary **T**est-**R**evised is *invalid* for measuring intelligence, but *valid* to measure receptive one-word vocabulary skills)

Types of Validity

1. Face validity
   - The extent to which a test is subjectively viewed as covering the concept it purports to measure

2. Content validity
   - *All test items* are relevant to measuring what *the test purports to measure*

3. Predictive validity
   - the *accuracy* with which a test predicts *future performance* of a client
   - (GRE test —> predicts academic performance)

4. Concurrent validity
   - the degree to which *a new test* correlates with an established *valid test*

5. Constructive validity
   - the degree to which *test scores* are consistent with *theoretical constructs or concepts*

Reliability

- The consistency or stability with which the same event is repeatedly measured. The assessment should give consistent results on repeated administrations

- Reliability of the test is **influenced** by several factors:
  - Fluctuations in examinee's behavior
  - Examiner's error (not following the procedures in the manual) - instrumentation or equipment errors

Types of Reliability

- **Test-Retest Reliability**
  - Consistency of measures when the same test is administered to the same person twice
- **Split-Half Reliability**
  - Refers to a test's internal consistency. Scores from one half of the test correlate with results from the other half of the test.
- **Rater Reliability: Inter-observer/ Inter-judge Reliability**
  - The more similarly the observers *independently* rate the same skill or event
- **Rater Reliability: Intra-observer/ Intra-jugde Reliability**
  - Consistency in which *the same observer* measures the same phenomenon on repeated occasions
- **Alternate Form Reliability (parallel form reliability)**
  - Consistency of measures when two forms of the same test are administered to the same person

Standardization

- Standardization is accomplished so that test-giver bias and other extraneous influences do not affect the client's performance and so that results from different people are comparable.

- Most of the standardized tests clinicians use are norm-referenced. But standardized is not synonymous with norm-referenced.

- Any type of test can be standardized as long as uniform test administration and scoring are used.

Standardization

- Test developers are responsible for clearly outlining the standardization and psychometric aspects of a test. Each test's manual should include information about:
  - The purpose(s) of the test
  - The age range for which the test is designed and standardized
  - Test construction and development
  - Administration and scoring procedures
  - The normative sample group and statistical information derived from it
  - Test reliability
  - Test validity

Standardized Test Administration

- Determining Chronological Age

- Basals and Ceilings

- Standardized Administration, Modification, and Accommodation

Determining Chronological Age

- Chronological age is the exact age of a person in years, months, and days. It is important for analyzing findings from standardized tests, as it allows the clinician to convert raw data into meaningful scores.

- To calculate chronological age:
    1. Record the test administration date as year, month, day.
    2. Record the client's birth date as year, month, day.
    3. Subtract the birth date from the test date.

Basals and Ceilings

- Basal refers to the starting point for test administration and scoring.

- Ceiling refers to the ending point.

- Basals and ceilings allow the tester to hone in on only the most relevant testing material. It would not be worthwhile or efficient, for example, to spend time assessing pre-speech babbling skills in a client who speaks in sentences, or vice versa.

Standardized Administration, Modification, and Accommodation

- Standardized tests are designed to be administered in a formulaic manner. That makes them, by definition, **standardized**. It is important to administer test items according to the protocol outlined in the test manual.

- **Accommodations** are minor adjustments to a testing situation that do not compromise a test's standardized procedure.

- **modifications** are changes to the test's standardized administration protocol. For example, a test giver might re-word or simplify instructions, allow extra time on timed tests, repeat prompts, offer verbal or visual cues, skip test items, allow the test taker to explain or correct responses, and so forth. Any such instance of altering the standardized manner of administration invalidates the norm-referenced scores.

Essential Reading for next time:

- Foundations of Assessment: The Code of Fair Testing Practices in Education (pages 16 -19)