- If you are viewing this course as a recorded course after the live webinar, you can use the scroll bar at the bottom of the player window to pause and navigate the course.
- This handout is for reference only. It may not include content identical to the PowerPoint. Any links included in the handout are current at the time of the live webinar, but are subject to change and may not be current at a later date.

ONLINE CONTINUING EDUCATION FOR THE LIFE OF YOUR CAREER

Interpreting Test Scores & Key Concepts with Standardized Tests

Carolyn (Carney) Sotto, PhD CCC-SLP

Moderated by:
Amy Natho, MS, CCC-SLP, CEU Administrator, SpeechPathology.com

Need assistance or technical support?

- Call 800-242-5183
- Email customerservice@SpeechPathology.com
- Use the Q&A pod

How to earn CEUs

- Must be logged in for full time requirement
- Log in to your account and go to Pending Courses
- Must pass 10-question multiple-choice exam with a score of **80%** or higher
  - Within **7 days** for live webinar; within **30 days** of registration for recorded/text/podcast formats
- Two opportunities to pass the exam

Interested in Volunteering to be a Peer Reviewer?

- APPLY TODAY!
- 3+ years SLP Professional Experience Required
- Contact Amy Natho at anatho@SpeechPathology.com

CONTINUED

Interpreting Test Scores & Key Concepts with Standardized Tests

Carolyn (Carney) Sotto, PhD CCC-SLP
University of Cincinnati
Department of Communication Sciences & Disorders
College of Allied Health Sciences
sottocd@ucmail.uc.edu

CONTINUED

Learning Outcomes

- As a result of this course, participants will be able to:
- 1) Discuss the standardization protocol of a standardized test
- 2) Identify the different types of scores found on standardized tests
- 3) Explain normal distribution and the bell-shaped curve

8

CONTINUED

Normal Curve Handout

- If you have access to a book or handout with the normal curve - please take it out to reference during this presentation

9

CONTINUED

Evidence-Based Practice

- ASHA's Joint Coordinating Committee on Evidence-Based Practice has a position statement (ASHA, 2005).
- One part part of the position statement in terms of this presentation is: For clinical practice to be evidence-based, speech-language pathologists & audiologists MUST:
  - evaluate prevention, screening, and diagnostic procedures, protocols and measures to identify maximally informative and cost-effective diagnostic and screening tools, using recognized appraisal criteria described in the evidence-based practice literature;

10

CONTINUED

Diagnostic Process

- Remember: standardized testing should be only one piece of information used in the evaluation process!
- Challenging Process
- Administering various tests & measurements
- Standardized Testing (norm-referenced & criterion-referenced)
- Interpreting Test Scores and making clinical decisions
- Case History
- Authentic Assessment
  - Observation
  - Interviews
  - Language Samples
  - School-work Samples

11

CONTINUED

Creating Effective Assessments

- Be thorough
- Use a variety of assessments (standardized, authentic, dynamic, etc.)
- Be sure your evaluation is valid.
- Be sure your evaluation is reliable.
- Individualize your assessment
- NO COOKBOOK assessment

(Shipley & McAfee, 2009)

12

Test Selection

- Select tests that meet the assessment's purpose
- Is the test appropriate for age level & age range for which it was standardized?
- Was the test well standardized (consider representativeness of samples, objectivity of administration and scoring, norms)
- Ask yourself this question: Is this test going to answer the question(s) you have about your client?

13

Test Selection (continued)

- Be familiar with the test (includes practicing administration of the test multiple times)
- Know the information you must have to make an informed decision
- Be a good consumer of tests
- Determine if you are qualified to administer the test

14

Psychometric Variables

- 3 areas of psychometrics to consider in test selection
  - Reliability
  - Validity
  - Standardization

15

Reliability

- **Reliability** refers to providing consistency with which a test gives results on repeated administrations or with different examiners evaluating the same test measurements.
- Reliability = Replicability
- Several Types of Reliability

16

CONTINUED

Types of Reliability

- **Several Different Types such as:**
  - Test/Retest, Alternative Form, Split-Half
- **Rater** - getting same results over multiple test administrations
  - Intra-rater - same person administers/scores same test
  - Inter-rater - different person administers or scores same test

17

CONTINUED

Validity

- **Validity** - refers to the extent a test measures what it claims to measure
- Validity = Truthfulness
- Validity - purpose test is intended for, may be valid for one purpose and invalid for another (Plante, 1996).
- Several Types of Validity

18

CONTINUED

Types of Validity

- Several Different types such as:
  - Content, Construct, Criterion Related (Concurrent & Predictive)

19

CONTINUED

Standardization

- AKA Formal tests or Norm-referenced tests
- Standard protocol followed for administration and scoring
- Results from different clinicians giving test are comparable
- All examiner's manuals should include:
  - Test purpose(s)
  - Test development and construction
  - Procedures for administration and scoring
  - Normative sample group and statistical information resulting from group
  - Reliability
  - Validity
- Read test reviews in peer reviewed journals

20

CONTINUED

Normative Sample

- Group of individuals tested in a given population and their data calculated statistically to allow comparisons between the person you are testing and the normative sample.
- All norm-referenced scores on a test come from this normative group
- Individual is given a test and his/her score is compared to performance of normative group

21

CONTINUED

How do we get Adequate Norms

- **Representative sample** - wide range of individuals having similar characteristics as those who will be taking the test. Variables such as age, gender, race, geographical area, and socioeconomic background
- **Number of persons included in the sample** - adequate numbers tested at each age level
- **Relevance of norms to purpose of the test** - usually using national norms based on general population

22

What do we need to know about Quantitative Data?

- We need to understand some basic concepts about quantitative data to correctly interpret test scores
- When tests are given or data are gathered from a large sample of people, the scores should form a "normal" or "bell-shaped" curve.
- Most scores cluster around the "average" or the "mean". The mean is known as a measure of central tendency.
- A few scores will be scattered away from the mean - either higher or lower

23

What else do we need to know about Quantitative Data? (cont.)

- Measure of Variability - AKA Standard Deviation
- Standard Deviation - the average difference of scores from the mean score
- Use of "Mean" & "Standard Deviation" helps both clinicians and researchers to know scoring patterns characteristic of the majority of a sampled population.
- The height and width of the bell curve determined by mean & standard deviation.

24

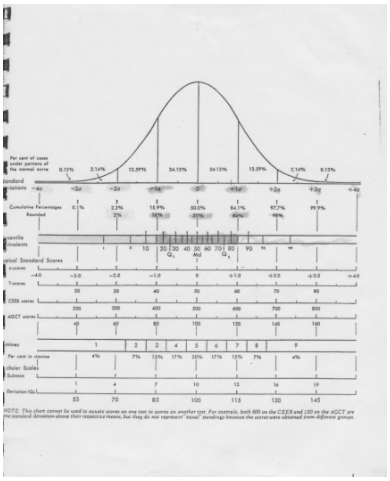
What should a Normal Distribution Look Like?

- Approximately 68% of the scores fall within 1 SD of the mean, 34% on each side
- Approximately 95% of scores fall within 2 SD of the mean, 47.5% on each side
- Approximately 99.7% of scores fall within 3 SD of the mean, 49.85% on each side
- Look at Bell Curve on next slide (good to use when reporting results to parents)

25

Normal Distribution Curve

- Darker blue (middle)=average
  -1 to +1 standard deviations
  [16th to 84th percentile ranks]
  Standard score of 85-115
- Yellow (left)=below average
  -2 to -4 standard deviations
  [2nd percentile rank and below]**
  Standard score of 70 and below
- Light blue (right)=above average
  +2 to +4 standard deviations
  [85th and above percentile ranks]
  Standard score of 130 and above



26

Types of Test Norms & Scores Reported

- Raw Scores
- Standard Scores
- Percentile Ranks
- Stanines
- Confidence Intervals
- Age & Grade Equivalents

27

Raw Scores

- Raw Scores - Raw scores by themselves are basically meaningless and should not be reported. They are needed to reference the important statistical scores in the test norms manual. Raw Scores are converted into some other type of derived score which has meaning

  Raw Scores cannot be compared and may represent something different according to each specific test, for example:

  1) the total number correct on a test

  2) the total number of errors on a test

28

CONTINUED

Standard Scores

- **Standard Score** indicates the distance the raw score is away from the mean or the average and can tell us how different an individual's score is from the average.
- Based on equal interval units
- Most preferred score for interpreting since based on the normal distribution
- Most standard scores we use as SLPs is the ***z-score.*** A z-distribution has a mean of 0 and a SD of 1.
- Reference point for a standard score is the individual's own age group at the time of testing
- SS are expressed in whole numbers and have a (see slide 19)
  - Mean - 100
  - Standard Deviation - 15
  - SS of 85 is 1 SD below mean
  - SS of 70 is 2 SD below mean

29

CONTINUED

Percentile Ranks

- Percentile Ranks - compares the individual's test performance with normative scores (or the percentage of people) that scored higher and lower than the given raw score.
  - Example: If a raw score of a student converts to the 18th percentile rank, this means that 18% of the normed sample earned similar or lower scores, and 82% of the population received higher scores.
- Percentile Rank and Percentage not the same
- Based on an ordinal or rank scale
- Mean or Average percentile rank is 50 (
- 1 Standard Deviation below mean = 16
- 2 Standard Deviations below mean = 2

30

Stanines

- Stanine (standard nine)= number to represent where the score is located (equally distributed) on curve.
- Another way to rank an individual's test performance
- A score based on a 9-unit scale, 5 is average performance or mean
- Note: most SLPs do not report stanines as frequently as standard scores and percentile ranks. Not on many of the tests that we use.

31

Age & Grade Equivalent Scores

- Obtained from norm-referenced tests
- Represented in years and months (e.g. 4 years, 3 months or 4-3 or 4:3 or 4;3)
- Relates the score client earned to an age or grade score
- A word of CAUTION:
  - On a Rank-Order Scale and not really comparing child's performance in relation to same peer group. Place child on a growth continuum which may or may not increase at regular intervals
  - Should only be used in combination with other test scores OR not used at all
  - Can be tempting to use b/c they appear to relate to development but they really do not!
  - Can be grossly misinterpreted by parents, physicians & educators (and yes, even SLPs)
  - Does not account for individual differences or the range of performances one might expect

32

Standard Error of Measurement

- ANY MEASURE is subject to error, especially human error.
- An estimate of how much a person's score is expected to vary in repeated administrations
- The **standard error of measurement** (SEM) was developed to increase the precision of whether the *observed score* (raw score/standard score) is reasonably close to the possible *true score* (score earned if no error existed - true score really does not exist).
- The SEM is the standard deviation of error around a person's true score and should be reported in the test manual for norm-referenced tests.
- When used along with the standard score, mean, standard deviation, SEM provides clinician with a powerful quantitative tool to describe and evaluate test performance.

33

Confidence Intervals

- The SEM is used to determine a **confidence interval**
- The range within which the "true" score falls
- Test Manuals often report the levels of confidence (68%; 90%; 95%)
- If given a choice, choose 90% level, since 90% level of significance is high enough with the smallest amount of measurement error.

34

Common Errors Using Norm-Referenced Tests

- Measuring Progress in Treatment
- Forgetting that Norm-Referenced tests distort what they are diagnosed to test
- Ignoring cultural groups making up the normative sample

35

Norm Reference vs Criterion Reference

- Majority of tests available
- Compare to sample that represents a population
- Always standardized & norm referenced
- How does the client compare to the group average? Emphasis on group similarity.

- Fewer tests available
- No comparison made - identifies strengths and weaknesses or specific levels of performance
- Standardized only for having standard procedures for administration and scoring
- How does the client compare to the expected performance? Emphasis on individual performance

36

Basals & Ceilings

- Standardized tests often have basals and ceilings for the administration of the test.
- A basal is the beginning point of the administration process when a specific number of items are scored as correct.
- A ceiling is the end point in the administration process when a specific number of items are scored as incorrect productions

37

Interpreting & Scoring Test Results

- Your manual is your guide!
- Be sure you are familiar with the examiner's manual.
- Be sure you have calculated the client's age correctly!
- Be certain you are looking at the correct age in the examiner's manual when scoring test results!
- Be aware that some tests have 2 forms.

38

Let's Practice

- If we have time.

39

References

- American Speech-Language-Hearing Association. (2005). Evidence-Based Practice in Communication Disorders [Position Statement]. Available from www.asha.org/policy.
- Haynes, W.O. & Pindzola, R.H. (2008). *Diagnosis and evaluation in speech pathology* (7th ed.). Boston, MA: Pearson Education Inc.
- Paul, R. & Cascella, P.W. (2007). *Introduction to clinical methods in communication disorders.*(2nd ed.). Baltimore, MD: Brookes Publishing Company.
- Plante, E. (1996). Observing and interpreting behaviors: An introduction to the clinical forum. *Language, Speech and Hearing Services in Schools, 27*, 99-101.
- Shipley, K.G. & McAfee, J.G. (2009). Assessment in speech-language pathology. (4th ed.). Clifton Park, NY: Delmar Cengage Learning.
- Wiig, E.H. & Secord, W.A. (1991). Measurement and assessment, A marriage worth saving. Lockport, NY: Educom Associates, Inc.

40