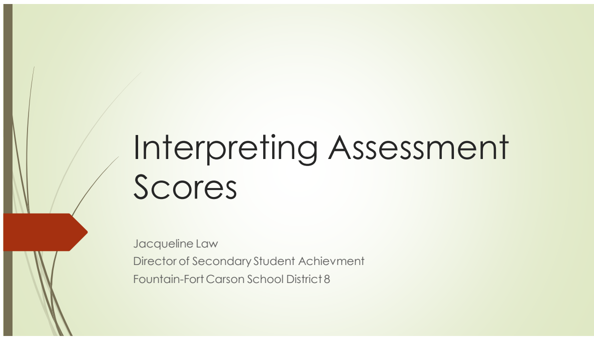
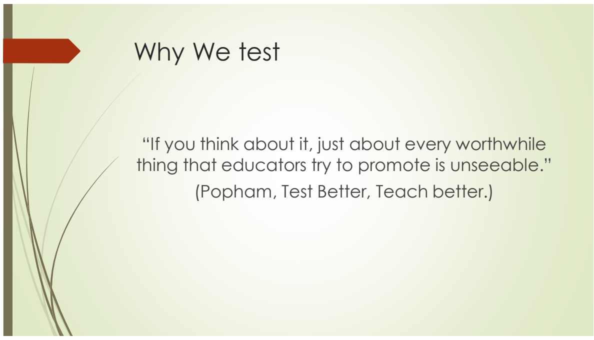


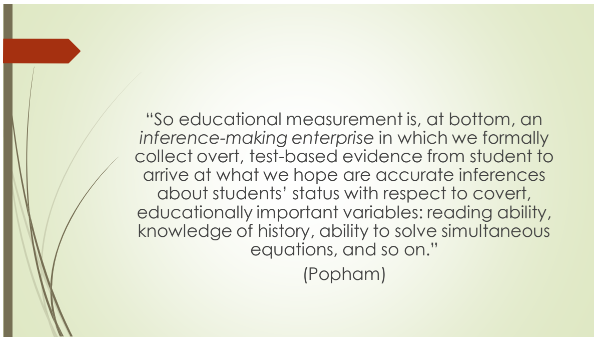
Interpreting Assessment Scores

Jacqueline Law
Director of Secondary Student Achievment
Fountain-Fort Carson School District 8

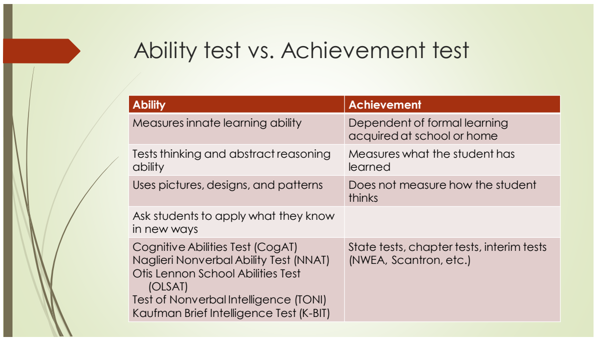


Why We test

"If you think about it, just about every worthwhile thing that educators try to promote is unseeable."
(Popham, Test Better, Teach better.)

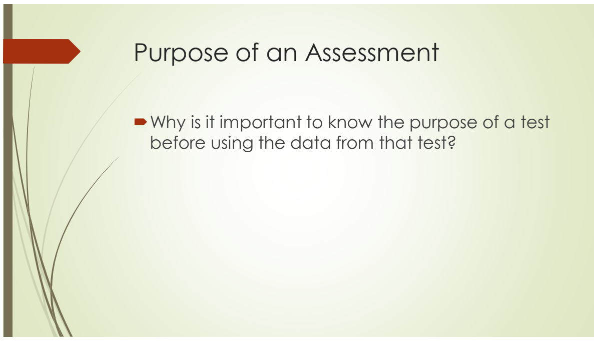


"So educational measurement is, at bottom, an *inference-making enterprise* in which we formally collect overt, test-based evidence from student to arrive at what we hope are accurate inferences about students' status with respect to covert, educationally important variables: reading ability, knowledge of history, ability to solve simultaneous equations, and so on."
(Popham)



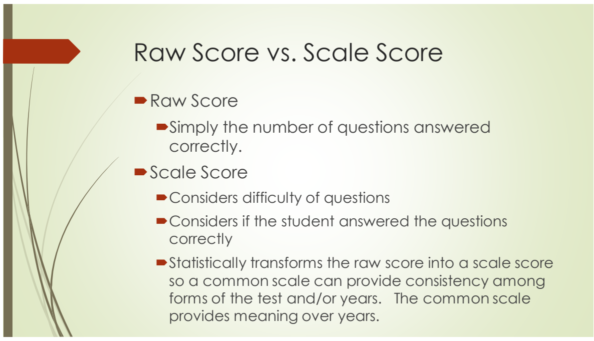
Ability test vs. Achievement test

Ability	Achievement
Measures innate learning ability	Dependent of formal learning acquired at school or home
Tests thinking and abstract reasoning ability	Measures what the student has learned
Uses pictures, designs, and patterns	Does not measure how the student thinks
Ask students to apply what they know in new ways	
Cognitive Abilities Test (CogAT) Naglieri Nonverbal Ability Test (NNAT) Otis Lennon School Abilities Test (OLSAT) Test of Nonverbal Intelligence (TONI) Kaufman Brief Intelligence Test (K-BIT)	State tests, chapter tests, interim tests (NWEA, Scantron, etc.)



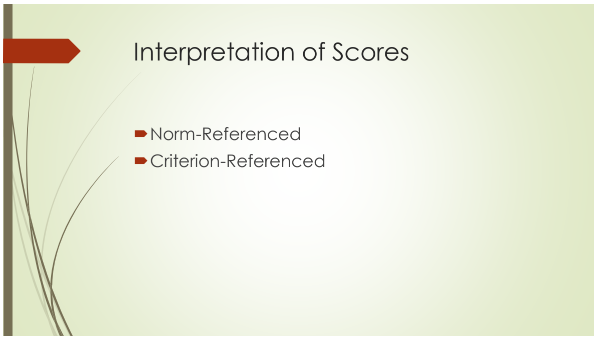
Purpose of an Assessment

- Why is it important to know the purpose of a test before using the data from that test?



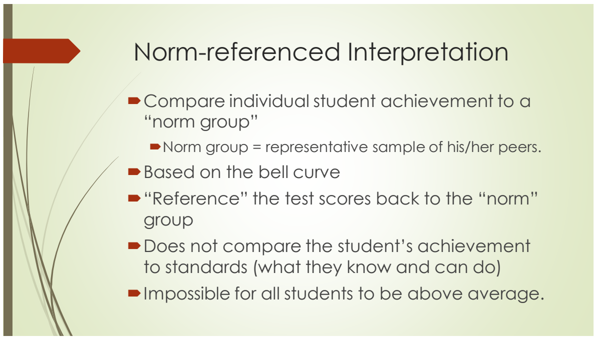
Raw Score vs. Scale Score

- Raw Score
  - Simply the number of questions answered correctly.
- Scale Score
  - Considers difficulty of questions
  - Considers if the student answered the questions correctly
  - Statistically transforms the raw score into a scale score so a common scale can provide consistency among forms of the test and/or years. The common scale provides meaning over years.



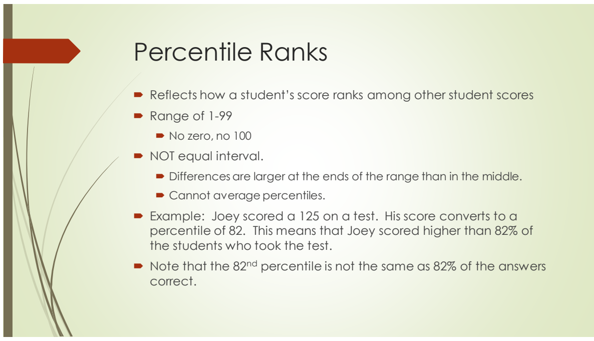
Interpretation of Scores

- Norm-Referenced
- Criterion-Referenced



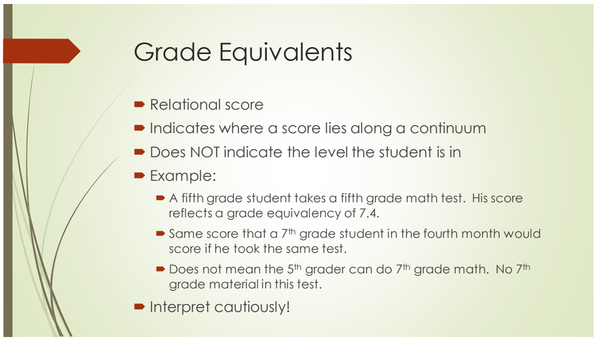
Norm-referenced Interpretation

- Compare individual student achievement to a "norm group"
  - Norm group = representative sample of his/her peers.
- Based on the bell curve
- "Reference" the test scores back to the "norm" group
- Does not compare the student's achievement to standards (what they know and can do)
- Impossible for all students to be above average.



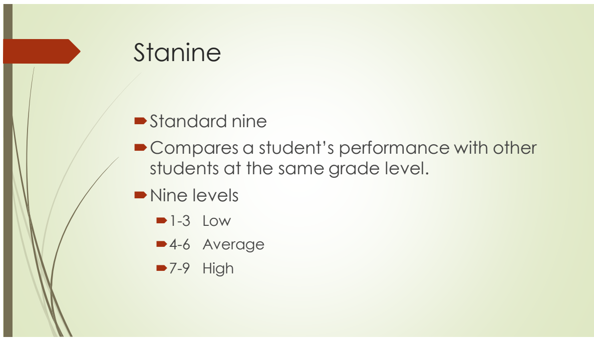
Percentile Ranks

- Reflects how a student's score ranks among other student scores
- Range of 1-99
  - No zero, no 100
- NOT equal interval.
  - Differences are larger at the ends of the range than in the middle.
  - Cannot average percentiles.
- Example: Joey scored a 125 on a test. His score converts to a percentile of 82. This means that Joey scored higher than 82% of the students who took the test.
- Note that the 82nd percentile is not the same as 82% of the answers correct.



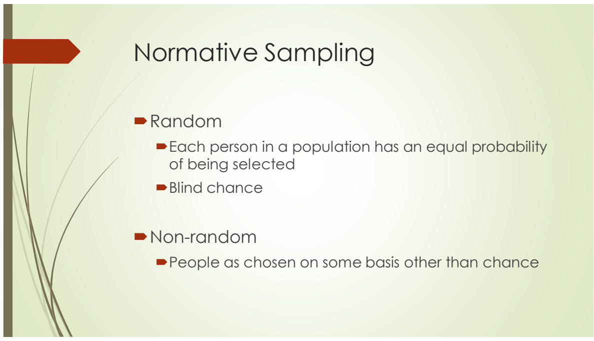
Grade Equivalents

- Relational score
- Indicates where a score lies along a continuum
- Does NOT indicate the level the student is in
- Example:
  - A fifth grade student takes a fifth grade math test. His score reflects a grade equivalency of 7.4.
  - Same score that a 7th grade student in the fourth month would score if he took the same test.
  - Does not mean the 5th grader can do 7th grade math. No 7th grade material in this test.
- Interpret cautiously!



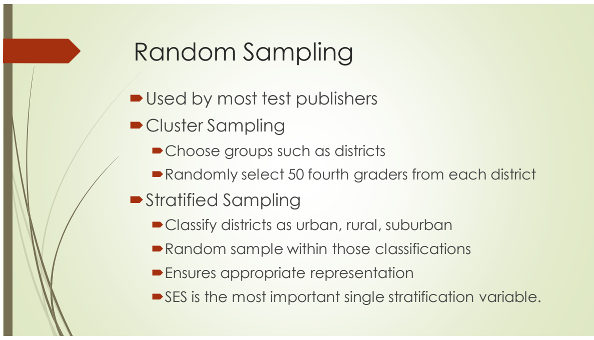
Stanine

- Standard nine
- Compares a student's performance with other students at the same grade level.
- Nine levels
  - 1-3 Low
  - 4-6 Average
  - 7-9 High



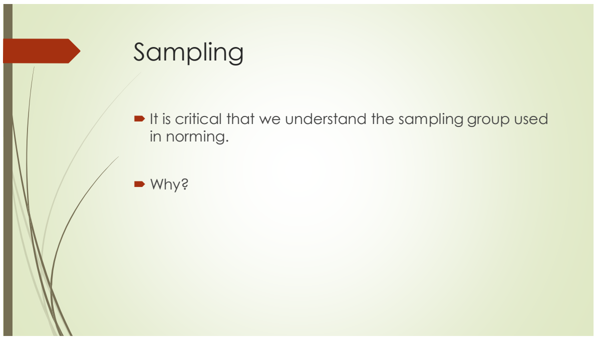
Normative Sampling

- Random
  - Each person in a population has an equal probability of being selected
  - Blind chance
- Non-random
  - People as chosen on some basis other than chance



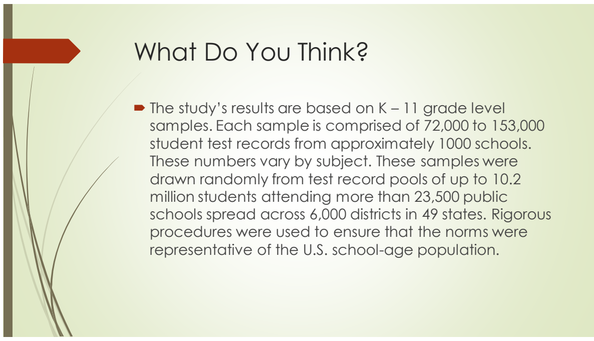
Random Sampling

- Used by most test publishers
- Cluster Sampling
  - Choose groups such as districts
  - Randomly select 50 fourth graders from each district
- Stratified Sampling
  - Classify districts as urban, rural, suburban
  - Random sample within those classifications
  - Ensures appropriate representation
  - SES is the most important single stratification variable.



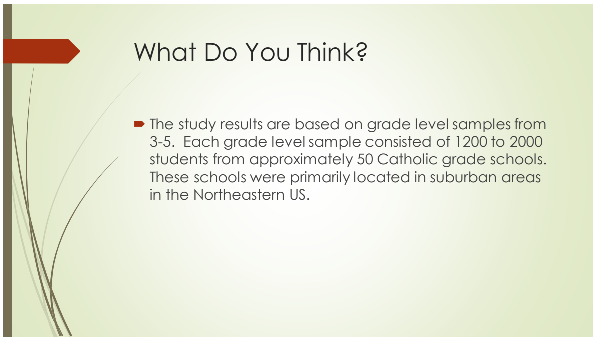
Sampling

- It is critical that we understand the sampling group used in norming.
- Why?



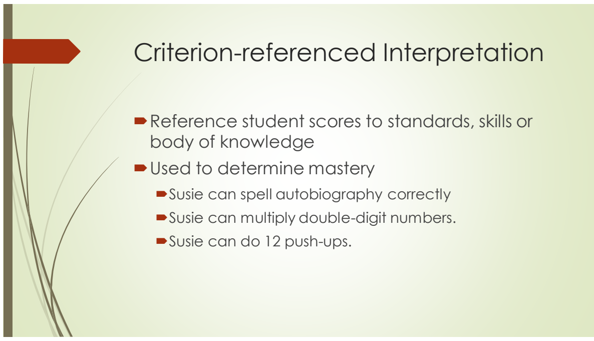
What Do You Think?

- The study's results are based on K – 11 grade level samples. Each sample is comprised of 72,000 to 153,000 student test records from approximately 1000 schools. These numbers vary by subject. These samples were drawn randomly from test record pools of up to 10.2 million students attending more than 23,500 public schools spread across 6,000 districts in 49 states. Rigorous procedures were used to ensure that the norms were representative of the U.S. school-age population.



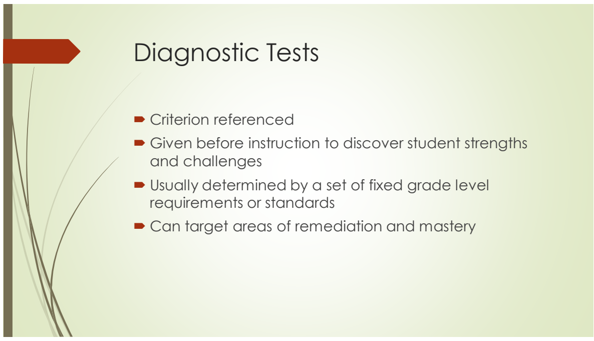
What Do You Think?

- The study results are based on grade level samples from 3-5. Each grade level sample consisted of 1200 to 2000 students from approximately 50 Catholic grade schools. These schools were primarily located in suburban areas in the Northeastern US.



Criterion-referenced Interpretation

- Reference student scores to standards, skills or body of knowledge
- Used to determine mastery
  - Susie can spell autobiography correctly
  - Susie can multiply double-digit numbers.
  - Susie can do 12 push-ups.

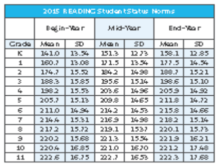


Diagnostic Tests

- Criterion referenced
- Given before instruction to discover student strengths and challenges
- Usually determined by a set of fixed grade level requirements or standards
- Can target areas of remediation and mastery

Standard Deviation

- In the status norms for Reading, grade 2 students in the middle of the "begin-year" period had a mean score of 174.7 and a standard deviation of 15.5. To get a sense of how much dispersion there was, the SD 15.5 can be subtracted from the mean and added to the mean to produce a range of about 159 –190. Since the norms are based on the bell curve, we know that 68% of all scores are expected to fall between in this range



2015 READING Student Status Norms

	Begin-Year		Mid-Year		End-Year	
Grade	Mean	SD	Mean	SD	Mean	SD
K	141.0	13.54	151.3	12.73	158.1	12.85
1	160.7	13.08	171.5	13.54	177.5	14.54
2	174.7	15.52	184.2	14.98	188.7	15.21
3	188.3	15.85	195.6	15.14	198.6	15.10
4	198.2	15.53	203.6	14.96	205.9	14.92
5	205.7	15.13	209.8	14.65	211.8	14.72
6	211.0	14.94	214.2	14.53	215.8	14.66
7	214.4	15.31	216.9	14.98	218.2	15.14
8	217.2	15.72	219.1	15.37	220.1	15.73
9	220.2	15.68	221.3	15.54	221.9	16.21
10	220.4	16.85	221.0	16.70	221.2	17.48
11	222.6	16.75	222.7	16.53	222.3	17.68

https://www.nwea.org/content/uploads/2015/06/2015-MAP-Normative-Data-AUG15.pdfur

Standard Error

- Variability of the mean (average)
- Estimation of how precise the score is
  - The smaller the standard error, the more precise the score.
- Mean plus or minus the standard error (SE or SEM)

Name	Score	SEM	Range	PR
Smith, Allison	319	3.2	316-322	67

Bias

- The test, or the interpretation or use of the results systematically disadvantages certain groups of students over others, such as students of color, students from lower-income backgrounds, students who are not proficient in the English language, or students who are not fluent in certain cultural customs and traditions. (edglossary.org)
- Cultural Bias
- Method Bias

Reliability

- Consistency
- Stability: Will a student taking the same test without instruction between those tests get the same results?
- Alternate forms: If Susie takes form A of a test, will she score as well on another form of the same test?
- Internal Consistency: Do items measuring the same construct produce similar scores?

Validity

- Does the test measure the student's learning relevant to a task or question, and nothing else?
- Example:
  - Does a word problem in math measure a student's reading as well as math ability?
  - If the question asks a student to write an answer to a social studies question, is the measurement about content or does it include assessing writing?

Resources

- http://edglossary.org/test-bias/
- https://www.nwea.org/resources/understanding-map-reports/
- Popham, W. James. *Test Better, Teach Better*. Alexandria, VA: ASCD, 2003. Print.
- "Bases for Assessing Normative Samples." *Student's Companion Website for Assessment for Effective Teaching*. Web. 11 Feb. 2016.
- "Interpreting Norm-Referenced Scores." CTB. Web. 11 Feb. 2016.