

Standardized Test Scores: Norm Referenced Scores

**To accompany Chapter 16,
*Developing and Using
Classroom Assessments*, 4th
ed. By Albert Oosterhof**

By Marlynn M. Griffin

All Rights Reserved, 2009



Before progressing any further...

- Complete the Introductory Terminology PPT
- Read pages 227-228 in your text; be sure you understand concepts related to all bold-faced words on these pages.



Norm-Referenced Test Scores

- Recall from Chapter 2 that scores can be referenced or compared to at least four benchmarks:
 - prior growth
 - ability
 - mastery of a well-defined content domain (criterion-referenced)
 - performance of others on the same measure (**norm-referenced**)
 - the focus of this chapter is norm-referenced scores



Interpreting Percentile Ranks

- One type of norm-referenced test score is a percentile rank.
- A percentile rank refers to the percentage of scores falling below the given score.
- If a student scores at the 85th percentile, that means that 85% of students scored below her score.



Interpreting Percentile Ranks

- Percentile ranks are expressed as whole numbers between 1 and 99
- They should be written as 85th percentile or 85%ile in order to avoid confusing them with a percent correct score.
- What's the difference between percent correct and a percentile score?
- Percent correct refers to what percent of items were answered correctly. If a student responds to 15 of 20 correctly, she has a score of 75% correct.



Interpreting Percentile Ranks

- Is 75% correct the same as the 75%ile? NO!
- 75% tells what proportion of items were answered correctly.
- 75%ile tells what percent of scores are below the given student's score.



Interpreting Percentile Ranks

- Let's look at our student who earned 15 of 20. We know that 15 of 20 is 75% correct.
- But is this an average score? Above average? Below average? – That is, when compared to the scores of other students, how can we classify this score?
- If many students answered fewer than 15 of 20 correct, then the score 15 may fall in a relatively high percentile – maybe as high as the 95thile, if answering 15 means the student outscored 95% of the group.

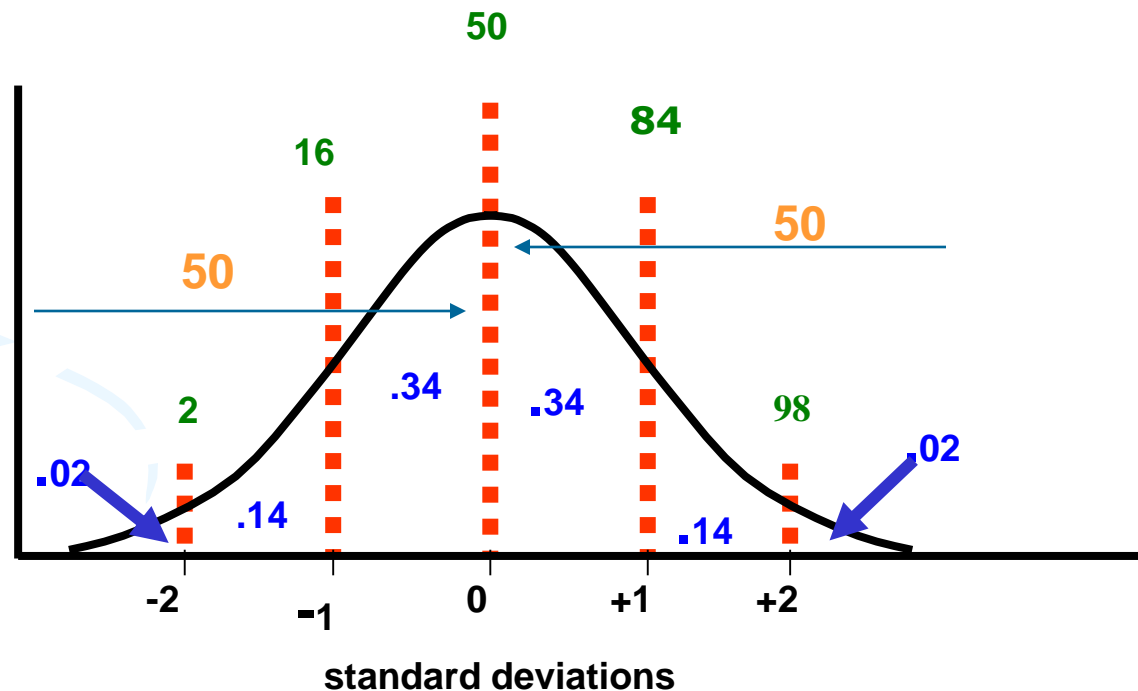


Interpreting Percentile Ranks

- Maybe 15 of 20 was one of the lowest scores in the group. If so, then this student could be in the 5%ile, or the 8%ile, or the 2%ile.
- If 15 is the average number of items correct for this group, then it may lead to a percentile ranking of something in the neighborhood of 50%ile.
- To simplify: the same raw score may lead to a high, low, average or somewhere in-between percentile rank **depending on the performance of the others in the group on the same measure.**

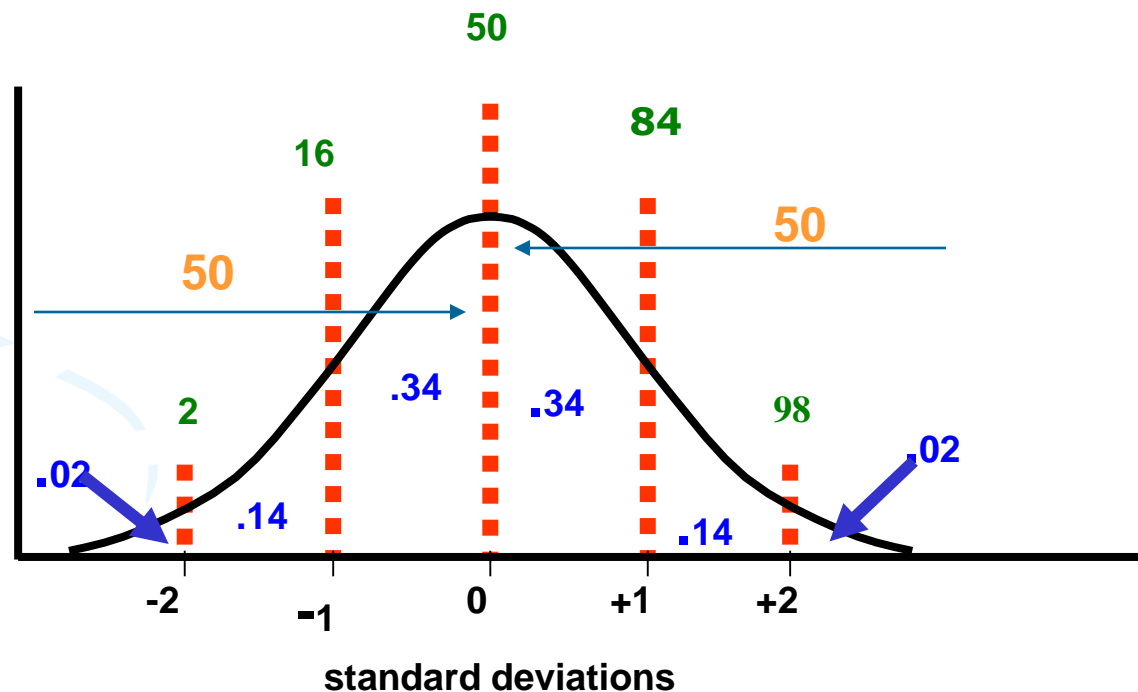
Interpreting Percentile Ranks

- Let's look back at the normal curve
- Note that if all percentages of areas under the curve on the left of the midline are added, they total 50
- Those to the right of midline also total 50



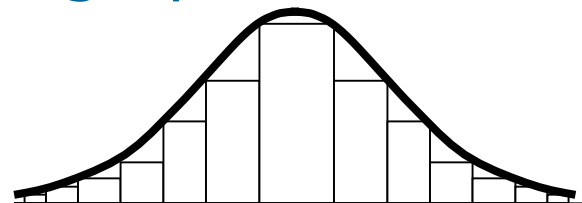
Interpreting Percentile Ranks

- Note also that the area under the curve at any given point is equal to the percentile rank at that point...WHAT?!?

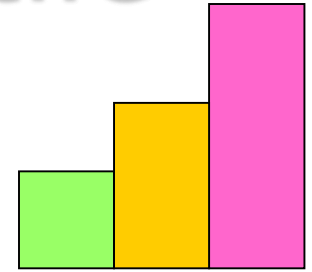


Interpreting Percentile Ranks

- Note also that the area under the curve at any given point is equal to the percentile rank at that point...WHAT?!?
- Look back at Slide 15 in the Introductory Terminology for Standardized Testing document. Recall that the normal curve represents a bar graph that displays how many people are in each bar, with more people represented by taller bars. You learned about bar graphs in elementary school; this is the same bar graph with an arched line connecting the bars.



Interpreting Percentile Ranks



- Look at the bar graph representing favorite color.
- 20% prefer green, 30% prefer orange, and 50% prefer pink.
- Moving from left to right, we can talk about cumulative percentages: 20% prefer green, 50% prefer green or orange, and 100% prefer green, orange, or pink.
- So, the bars added together represent the favorite colors of 100% of the population.

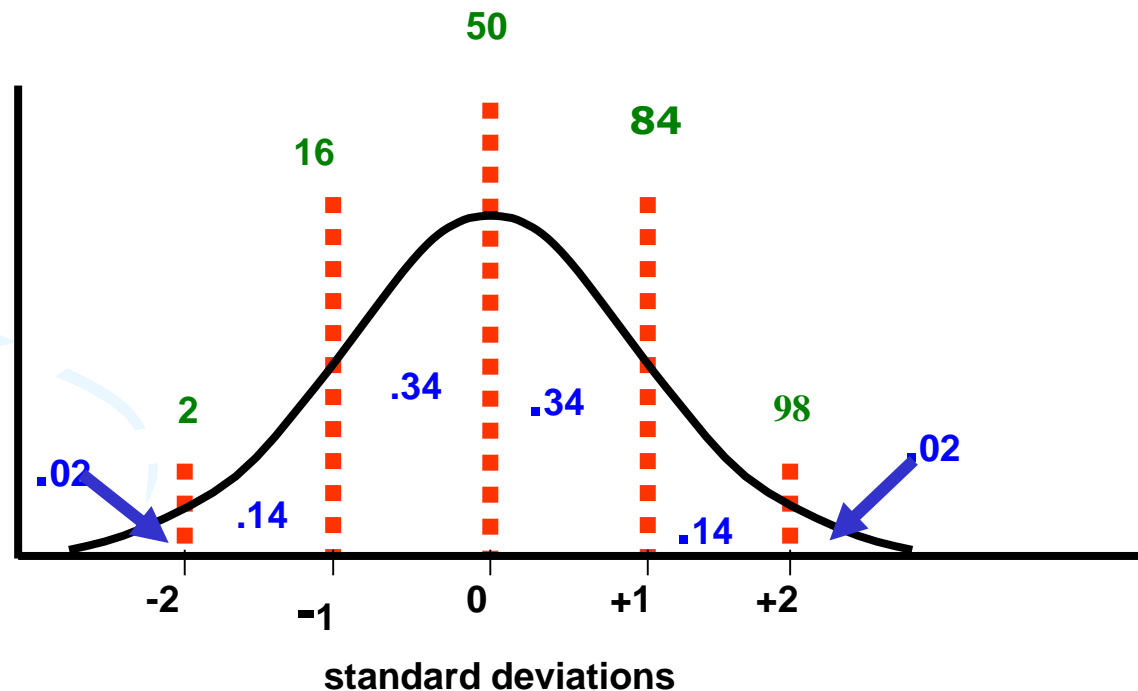


Interpreting Percentile Ranks

- The same thing applies to the normal curve.
- The first space on the left is 2% of the population, the second is 14%, the third is 34%. Crossing midline, the first space is 34%, the second is 14% and the third is 2%.
- Cumulatively, then, the first space is 2% of the population, the first and second are 2% plus 14% or 16% total, and adding the third brings us to midline at $16 + 34 = 50\%$ of the population.

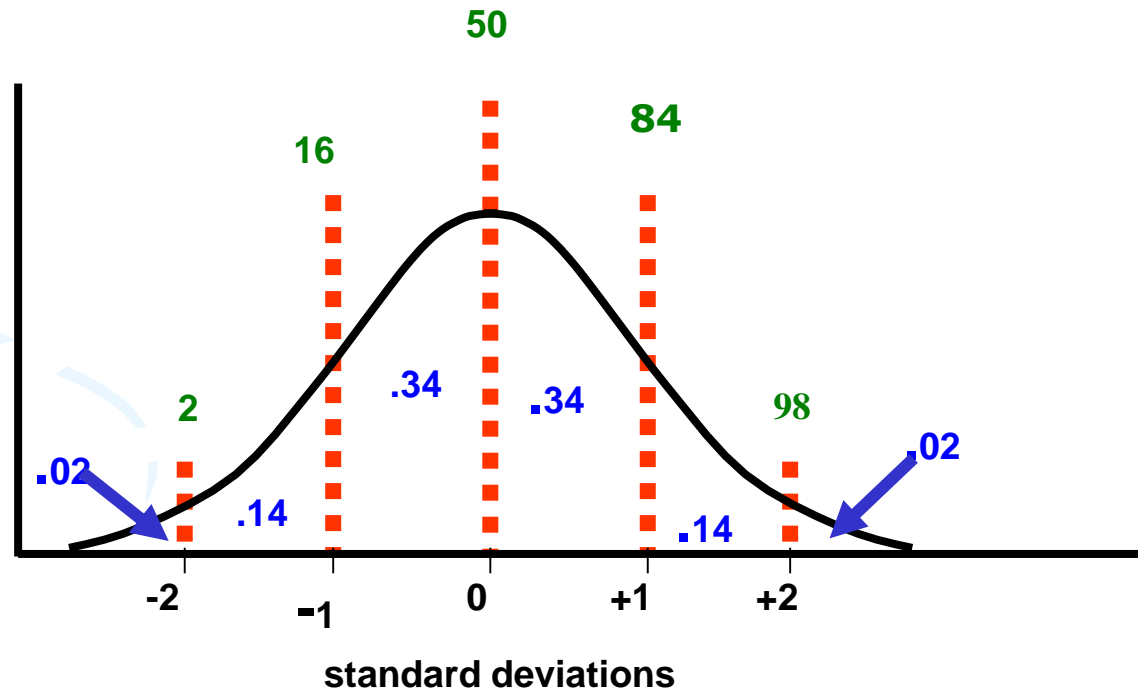
Interpreting Percentile Ranks

- Moving across the midline, 34% is added to the cumulative totals so that the first line to the right of midline represents 84% of the population; the second line to the right is $84 + 14 = 98\%$ and the end of the graph is $98 + 2\% = 100\%$ of the population.



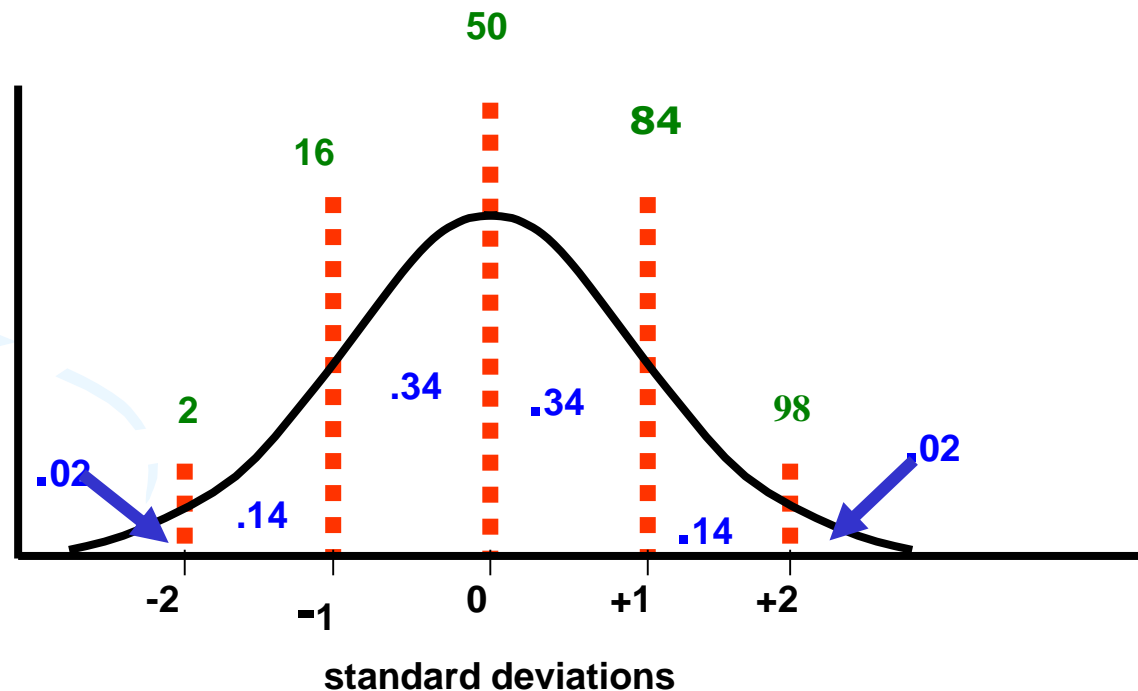
Interpreting Percentile Ranks

- Percentile ranks map onto this curve in much the same way. If you obtain the mean score on a measure (say, you are average height) then you will be at the 50%ile – your height will be greater than that of 50% of the people measured and your height would fall on the midline of the normal curve, as midline represents mean score and 50%ile.



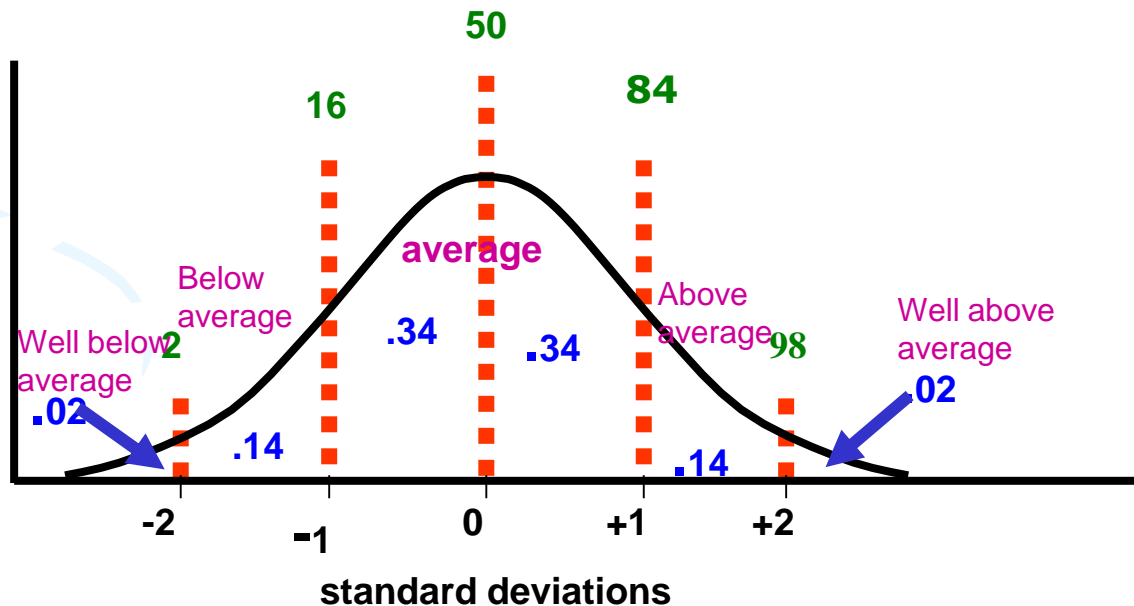
Interpreting Percentile Ranks

- Note also that the percentile rank of 50 places you at the mean score, which is 0 standard deviations from the mean score. A percentile rank of 84 places you 1 standard deviation from the mean (see the +1 at the bottom of the curve that represents 1 SD above the mean).



Interpreting Percentile Ranks

- Typically, all scores between 1SD below and 1SD above the mean (between the 16%ile and 84%ile) are considered to be in the average range.
- Those between 1SD above and 2SD above are above average, and between 2SD and 3 SD above are well above average. By the same token, between 1SD and 2 SD below is below average, and 2SD to 3SD below is well below average.





Interpreting Percentile Ranks

- **Percentile ranks, in sum:**
 - Report performance of an individual in comparison to the performance of a group.
 - Describe what percent of scores fall below the given percentile score.
 - Can be “mapped” on the normal curve to provide a visual representation of meaning.



Practice Interpreting Percentile Ranks

- John is in the 65th percentile on reading. This means that John scored better than _____ percent of others on reading.
- What percent of students scored higher than John on the reading test?
- Would John's reading score be considered average? Below average? Above average?
- What is John's score on the reading test?



Practice Interpreting Percentile Ranks

- John is in the 65%ile on reading. This means that John scored better than **65** percent of others on reading.
- What percent of students scored higher than John on the reading test? **35**
- Would John's reading score be considered **average**? Below average? Above average?
- What is John's score on the reading test? **Impossible to tell. Norm-referenced scores, such as percentile ranks, do not tell the number or percent correct on a test, just the student's standing in comparison with others.**



Practice Interpreting Percentile Ranks

- Jane scored in the 98%ile on math computation. How many standard deviations above the mean was Jane's score?
- Jane's score of 98%ile can be considered what? Average? Well above average? Above average?
- How can Jane's score on math computation be explained so her parents understand it?



Practice Interpreting Percentile Ranks

- Jane scored in the 98%ile on math computation. How many standard deviations above the mean was Jane's score? 2 (look at the curve; 98%ile falls on the same line as +2SD)
- Jane's score of 98%ile can be considered what? Average? Well above average? Above average?
- How can Jane's score on math computation be explained so her parents understand it? Jane's score is well above average. By that, I mean that if 100 students took this math test, Jane's score would be higher than the score of 98 of those students.

A decorative graphic on the left side of the slide featuring three balloons: a light green one at the top, a light blue one in the middle, and a light purple one at the bottom. Each balloon is attached to a streamer and has several small yellow triangular flags hanging from it.

Take a Break!

- If you've been viewing this presentation all in one sitting, you need a break now!
- Go get something to drink and get a dessert – you deserve a treat.



Interpreting Stanines

- Value of stanine is limited to whole numbers from 1 to 9
- The mean stanine score is 5 and the SD is 2
- Since stanines range only from 1 – 9, any student whose score is more than 2SD above the mean is assigned a score of 9, and more than 2SD below the mean is assigned a stanine score of 1



Interpreting Stanines

- Stanine scores can be represented as ranges of percentile ranks as follows:

a stanine of 1 = 0-3%ile

2 = 4-11%ile

3 = 12-23%ile

4 = 24-39%ile

5 = 40-59%ile

6 = 60-75%ile

7 = 76-87%ile

8 = 88-95%ile

9 = 96-99%ile



Interpreting Stanines

- Since stanine scores represent a range of percentile ranks, they take into consideration the possibility of measurement error. (Recall discussion of measurement error from the Reliability and Validity ppt, slide 7, in Module 1.)
- Remember that any given score is actually a representation of a student's true score (what they really, truly know) plus or minus other factors that influence their score on the test (plus: guessed correctly; clue to an answer in another question) (minus: feel sick on day of test; too noisy in hall during test)



Interpreting Stanines

- Thus, a student with a stanine of 7 is considered to be somewhere between the 76%ile and the 87%ile. She has outscored between 76 and 87% of students on this measure.
- While this may not seem as accurate as a percentile rank it may, in fact, be a bit more accurate because it implicitly acknowledges the existence of measurement error.



Interpreting Stanines

- So, back to the student with the stanine score of 7. We may find that her percentile rank is given as the 80%ile.
- Combined with her stanine score, we could say that the best estimate of her performance is that her score was higher than that of 80% of the test-takers, but it could, in fact be anywhere between the 76%ile and the 87%ile.



Interpreting Grade Equivalent Scores (GE)

- Grade equivalent scores convey growth. (see text, p. 237 for more explanation)
- Neither standard scores nor percentile ranks convey growth.
- Grade equivalents are computed to one place beyond the decimal point, but this last digit does not convey months of the school year.



Interpreting Grade Equivalent Scores (GE)

- Before progressing any further in this slide show, read pages 238 – 240 (stop in middle of column 1 at *Grade Equivalents Identify Median Scores*)
- This section describes how grade equivalent scores are established. For a brief summary, see the next slide.

A decorative graphic on the left side of the slide features three balloons in light green, light blue, and light purple, each with yellow streamers trailing behind them.

Interpreting Grade Equivalent Scores (GE)

- For example, a fourth grade mathematics test is administered to students in the 3rd, 4th, and 5th grades.
- The median raw score for each grade level is computed, and is set at 4.0 for 4th grade, 3.0 for 3rd, and 5.0 for 5th.
- Grade equivalents are then established for scores falling between the observed medians, such as 4.2, 4.5. 4.7.

A decorative graphic on the left side of the slide featuring three balloons: a light green one at the top, a light blue one in the middle, and a light purple one at the bottom. Each balloon has a streamer and is surrounded by several small yellow triangular shapes, resembling confetti or streamer tassels.

Interpreting Grade Equivalent Scores (GE)

- How are these scores interpreted?
- NOT with a direct interpretation. That is, it is erroneous to say that a 4th grader who earned a GE of 7.5 on a reading test is reading at the 7th grade level.



Interpreting Grade Equivalent Scores (GE)

- Rather, we can say that this 4th grader earned a score on the 4th grade reading test that represents the median (middle) score on this test for students in the 7th grade IF students in the 7th grade had taken the test.
- Remember, this 4th grade test is given to students in 3rd, 4th, and 5th grades. No 7th graders took it; GE scores of 7.0 were extrapolated from the data.

A decorative graphic on the left side of the slide featuring three balloons: a light green one at the top, a light blue one in the middle, and a light purple one at the bottom. Yellow streamers and triangular flags are scattered around the balloons.

Interpreting Grade Equivalent Scores (GE)

- This is an important distinction to make because many parents think they understand GE scores, and are thrilled when their little 2nd grade Susie is reading at the 7th grade level.
- But, would Susie really be interested in 7th grade books, with many words and no pictures?
- Would Susie's parents want her reading books with this more mature content?



Interpreting Grade Equivalent Scores (GE)

- We don't know at what grade level she is reading – we know only that she scored above the median on this test of reading comprehension, and also scored higher than students obtaining a score of 5.0 or 6.3 or 6.8.
- That is, GE scores indicate relative performance – how well a student performs when compared to other students.



ITBS Score Report

- The following slide shows a portion of an ITBS score report.
- The small insert on the upper part of the slide shows that the student was in second grade when this test was completed.
- The larger box provides a summary of student performance.

ITBS Score Report

Student ID:

Form/Level: A/ 8

















Test Date: 03/2006

Norms: Spring 2000


Order No.: 000179971

Page: 2

Grade: 2

Tests	SS	Scores GE	NPR	PERCENTILE RANK				High 99
				Low 1	25	50	75	
Vocabulary	170	2.9	54					
Reading Comprehension	193	4.3	87					
Reading Total	182	3.6	74					
Word Analysis	203	5.0	89					
Listening	171	2.9	58					
Spelling	186	3.9	89					
**Language	180	3.4	74					
Math Concepts	166	2.7	45					
Math Problems	176	3.2	66					
Math Computation	180	3.6	81					
Mathematics Total	174	3.2	64					
CORE TOTAL	179	3.3	72					
Social Studies	156	2.1	25					
Science	187	3.9	81					
Sources of Information	186	3.8	85					
COMPOSITE	180	3.4	75					

Legend: SS = Standard Score
GE = Grade Equivalent
NPR = National Percentile Rank

 National Percentile Rank




ITBS Score Report

- Note the following types of scores reported on this report:
- SS = standard scores (not discussed in this class)
- GE = grade equivalents
- NPR = national percentile rank (simply means the percentile rank is calculated on a national sample, or students from across the nation)
- Percentile rank illustrated by a bar graph



ITBS Score Report

- The ITBS reports performance on a variety of subtests. This performance is reported as subtest scores, scores on domains within subtests, and composite scores.
 - For example, the subtest of Reading is comprised of the areas Vocabulary and Reading Comprehension.
 - The composite Core score is comprised of combined student performance on the Reading, Language and Mathematics subtests.
 - The Composite score is a combination of all subtests on the ITBS.
- 



ITBS Score Report

- The ITBS uses a slightly different average range than what is considered average on the normal curve.
- On the normal curve, anything between the 16th and 84th %iles is considered average.
- On the ITBS, average is considered to fall between the 25th and 75th %iles.



ITBS Score Report

- Note that none of the student's scores fall in the below average range.
- Thus, any areas of weakness are areas of relative weakness for this child.
- That is, there is no below average performance, so "low" scores are those considered low for this child.



ITBS Score Report

- What are the child's areas of relative weakness?

*Note: relative weakness means compared to other areas tested, where is this child weak. The child may have NO areas that are below average, but if all areas are above average and one is lower than the others, though still above average, then it is a "relative" weakness. The same principle applies for relative strength.

- What are the child's areas of relative strength?



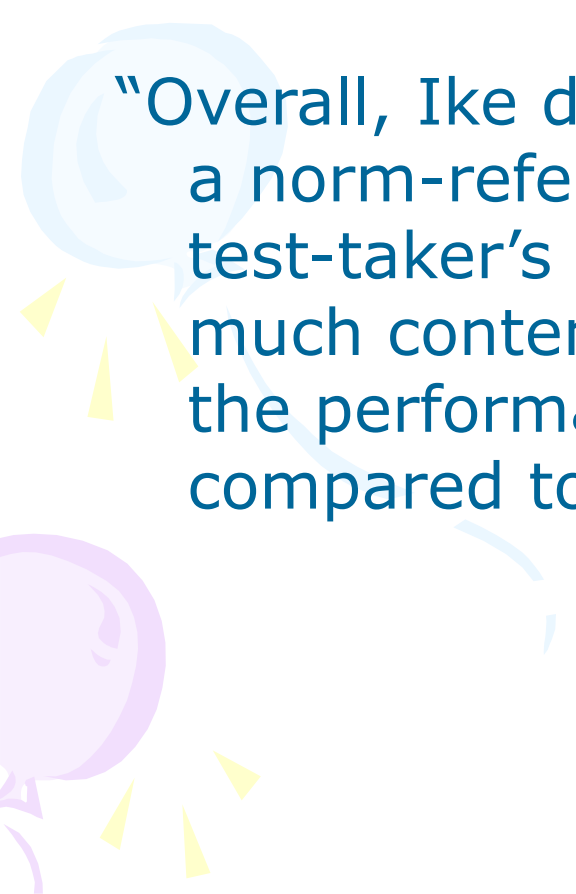
ITBS Score Report

- What are the child's areas of relative weakness?
Vocabulary, listening, math concepts, and social studies
- What are the child's areas of relative strength?
Reading comprehension, word analysis, spelling, math computation, science, sources of information



ITBS Score Report

If you wanted to offer a summary explanation of these scores to a parent, it might look like this:



“Overall, Ike did quite well on this ITBS. The ITBS is a norm-referenced test which means that the test-taker’s scores do not directly reflect how much content was mastered, but rather reflect the performance of the given student when compared to the performance of other students.”



ITBS Score Report

"Ike scored in the average or above average range on all components of this test. He scored highest on the subtests reading comprehension, word analysis, spelling, math computation, science and sources of information. His lowest scores were in the areas of vocabulary, math concepts, and social studies. These lower scores are not really "low" scores, they are just relatively low. That is, they are low for Ike but are still in the range of average performance."



Take a Break!

- This is the end of the Norm-Referenced Test Scores presentation.
- The next slide begins a presentation on Criterion-Referenced Test Scores.
- Following that presentation is a presentation entitled "Putting It All Together" that describes how to integrate data from a variety of sources when talking with parents.



Criterion Referenced Test Scores

**To accompany Chapter 16,
*Developing and Using
Classroom Assessments*, 4th
ed. By Albert Oosterhof**

By Marlynn M. Griffin

All Rights Reserved, 2009



Georgia CRCT

- The Georgia Criterion-Referenced Competency Tests (CRCT) are a set of tests given each year as part of the state's Adequate Yearly Progress (AYP) indicator



Georgia CRCT

- As criterion-referenced tests, the critical information from the score reports for these tests is what level the student obtained on the test
- There are 3 possible performance levels on this test
 - Level 1 or does not meet standard
 - Level 2 or meets standard
 - Level 3 or exceeds standard



Georgia CRCT

- Note that as a criterion-referenced test, student performance is not being compared to the performance of other students, but is compared to a pre-set standard that indicates mastery of content (levels 2 & 3) or non-mastery of content (level 1)



Georgia CRCT

- Scores are reported as scale scores. It is beyond the scope of this presentation to determine how scale scores are obtained.
- Scales with different ranges of scores can be set to be equivalent to each other, even though the score ranges are different.



Georgia CRCT

- For example, CRCT reports based on the Georgia Quality Core Curriculum (QCC) used a scale on which scores between 300 and 350 were considered Level 2, or meets standard, scores.
- CRCT reports based on the Georgia Performance Standards (GPS) use a scale on which Level 2 scores fall between 800 and 850.
- Thus, scores between 300 and 350 on one scale for the CRCT are equivalent to scores between 800 and 850 on a different scale for the CRCT.



Georgia CRCT

- When interpreting student performance on the CRCT, one simply looks at where a student's score falls in comparison to the Level 2 (meets standard) standard score range.
- Look at the following slide for a sample score report. You may want to print this slide so you have it handy.



Georgia CRCT

- I'll be referring to the handwritten numbers on the report as I discuss it.
- 1 & 2 describe the test and test taker, respectively. Note that this test from 2006 reports some scores based on GPS and one based on QCC. Current reports will show GPS only, but eventually the curriculum will change and we'll have transitional testing years again.

Georgia CRCT





Georgia CRCT

- 3a – 3e are column headings for the information listed on the report.
 - “a” column tells the test or subtest being described and the number of items in the test or subtest. Note that the Reading Test is comprised of 40 items, made up of 8 items on the vocabulary subtest and 32 on the comprehension subtest.

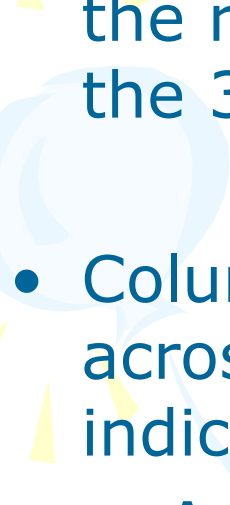



Georgia CRCT

- Column 3b indicates the student performance level. This is reported as 1, 2, or 3, with these numbers indicating whether the student is below standard, meets standard, or exceeds standard expectations.
- Column 3c reports the student's actual numerical scale score at the top and the number of items correct in each subtest below that.
 - Note that the student's scale score and performance level provide the same information but in different forms (whether or not the student met the standard) – if the student is at Level 2, their scale score is between 800 and 850; Level 1, below 800 and Level 3, above 850.



Georgia CRCT

- Column 3d indicates that the scale or GPS appears above the scale for QCC (follow the arrow to see the actual numbers). The scales are aligned on the report to show the correspondence between the 300 and 800 scales.
 - Column 3e provides the performance level cutoffs across the top, and below that the cross marks indicate the student's performance.
 - Again, these marks are consistent with the information provided in columns 3b and 3c, but are reported in slightly different form.
- 
- 



Georgia CRCT

- In short, here's the critical information about interpreting CRCT scores:
 - Look at the 3e column and see where the cross marks fall.
 - Are the marks in the does not meet column? Meets? Exceeds?
 - On this report, the student performed at level 2 for Reading and English/Language Arts, and at Level 3 for Mathematics.
 - So, he met standard for Reading and ELA, and exceeded in Mathematics.




Georgia CRCT

- Print the CRCT report on the next slide and use it to respond to the following.
1. Offer a brief summary of student performance in all three areas; that is, at what level did the student perform in reading, ELA, and math?
 2. What is the student's scale score in each of these areas?
 3. Why is the math score so much lower than the other two scores?
 4. How many items were on the Geometry and Measurement subtest? How many items did the student answer correctly?
 5. How many items on Reading for Information? How many answered correctly?



Georgia CRCT

1. Offer a brief summary of student performance in all three areas; that is, at what level did the student perform in reading, ELA, and math? **Student did not meet standard in any areas; performed at Level 1.**
 2. What is the student's scale score in each of these areas?
Reading – 792; ELA – 779; Math - 290
 3. Why is the math score so much lower than the other two scores? **The math appears lower because it is on a different scale, but it is in fact very similar to the other scores, in that it is in Level 1, slightly below the Level cut off.**
 4. How many items were on the Geometry and Measurement subtest? How many items did the student answer correctly?
14 items, 9 correct
 5. How many items on Reading for Information? How many answered correctly? **8 items, 5 correct**
- 

The background features abstract, colorful swirls in shades of purple, green, and blue, interspersed with small yellow triangles. The main title is centered in a large, bold, purple font.

Putting It All Together

**To accompany Chapter 16,
*Developing and Using
Classroom Assessments*, 4th
ed. By Albert Oosterhof**

By Marlynn M. Griffin

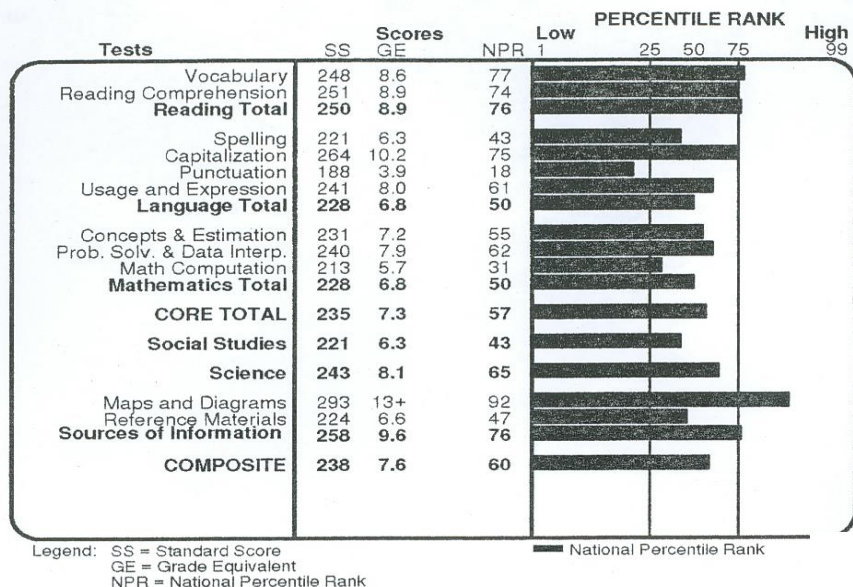
All Rights Reserved, 2009



Putting It All Together

- Student performance data will be gathered on CRCT tests, ITBS tests, daily performance, and through grading period averages.
- Teachers need to be able to integrate data from all of these sources to communicate accurately to parents about student performance.
- The Department of Education and school administrators tend to place a greater emphasis on CRCT scores than other sources of information when examining student performance. However, it's important to recognize that all data contribute different information about a student's performance and that patterns of performance need to be considered.

Partial ITBS Report



Putting it All Together

Partial CRCT Report (same type of information as on previous slide, but presented in a different format. This format is also presented on individual student score reports)

Student Grades

Reading A
Language Arts B
Social Science C
Science B
Mathematics B

Content Area	Scale Score	Performance Levels			
		Level 1 Does Not Meet	Level 2 Meets	Level 3 Exceeds	
			800 300	850 350	
Reading – GPS	856				
English/Language Arts – GPS	810				
Mathematics – QCC	327				



Putting It All Together

- Looking at all data presented on the previous slide, we can make the following summary statement about student performance. (print previous slide to examine while reading explanation on next slide)
- This statement is made as though talking to parents about performance.
- Note that the instruments used to collect the various types of data are explained, as well as the scores collected with those tools.



Putting It All Together

- Hi Mrs. Jones. I'm glad you came in to talk about Johnny's recent test scores. We have information from two different tests as well as his class grade averages. All of this information is useful in helping us determine how Johnny's doing. The two tests we use to examine performance are the Iowa Tests of Basic Skills (ITBS) and the Georgia Criterion-Referenced Competency Tests (CRCT).
- The ITBS is a norm-referenced test, which means that Johnny's score is reported in terms of how well he performs in comparison with others his age who have taken this same test.



Putting It All Together

- The CRCT is a criterion-referenced test, which means that Johnny's score is interpreted in terms of how well he has performed in comparison to the score on this test that indicates mastery.
- On the CRCT, Johnny met the standard on both English/Language Arts and Mathematics, and exceeded the standard in Reading. His ITBS scores are consistent with this performance. In ELA and in math, Johnny performed at the 50th percentile, which means he scored as well as or better than 50% of the kids who took this test.



Putting It All Together

- The 50th percentile indicates that his scores are in the average range, which is consistent with meeting the standard on the CRCT. His grades in ELA and in math are B. Typically, grades of B and C indicate are in the average range of performance for kids, so his grades are consistent with his test scores.
- While Johnny did well on all portions of the CRCT and the ITBS, his strong subject is reading. As you can see, he exceeded the standard on the CRCT in reading, he has a class grade of "A" in reading, and he is in the 76th percentile on the ITBS, indicating he performed better than 76 percent of other kids on the reading subtest.

Additional Georgia CRCT Information

- There is an 80 page online-guide to interpreting various CRCT score report forms. You don't need to spend time with this report for this class, but if you need help with interpretation in the future, there are plenty of resources in addition to this one comprehensive resource.
- <http://www.doe.k12.ga.us/DMGetDocument.aspx/CRCT%20Score%20Interpretation%20Guide%202008.pdf?p=6CC6799F8C1371F6BE7766F81D6BA67382079D4AE4D8E544B96E134DBFB9A428&Type=D>
- The Georgia CRCT report will begin reporting Lexile scores for Reading and English/Language Arts. While it's beyond the scope of this class to look at this information, the link below will take you to a powerpoint presentation that will help you to understand what you need to about Lexiles to be better prepared professionally.
- www.fultonschools.org/dept/langarts/teachersinfo/The%20Lexile%20Framework%20GRA%2006.ppt



That's All!

- Whew! This was difficult information to present in a slide show, with no live interaction.
- I know it was difficult information for you to process, too, and it was a BUNCH of information.
- Make sure that you completed all practice activities and that you understand the content thoroughly.
- You will use this information yearly as you interpret test results and communicate with parents.
- Go do something relaxing for a little while! You need a break!