

Understanding Reading Test Scores

There are a number of important reasons why teachers use tests (see the section titled 'The Purpose of Assessment'). Teachers are under some pressure to provide information to school authorities and parents about children's progress and attainment in reading; it is important to know to what degree the pupil is absorbing what the teacher is trying to teach; and it is important to compare the pupil's progress and level of attainment with that of other pupils of a similar age or class level.

A standardised, objective reading test is one means of determining with some precision the extent to which a pupil has approached one or more of the goals of a school's reading instruction programme. A reading test can help the test giver determine whether or not the subject's reading skills are as well developed as others of her or his age. Reading tests are also useful to monitor progress over time.

Some reading tests facilitate monitoring by enabling teachers to convert the raw score on the test to a reading age. To the novice test giver, the concept of reading age seems simple, e.g. an eleven-year-old should have a reading age of eleven and, therefore if the raw score converts to a reading age of nine, he or she is two years behind in reading and therefore has a reading problem. Unfortunately, it is not quite that simple. While describing reading ability in terms of reading age is very common, it is the most ambiguous and misleading method of interpreting reading test performance. This is mainly due to the fact that there is a variety of ways used by test designers to derive reading age scales.

The main problem with reading age tends to be a lack of understanding, which leads to a tendency to invest reading ages with a meaning and authority which is out of all proportions to their statistical origins. Their reference to age seems to imply something about the development of reading, as if certain skills and abilities were associated with particular reading ages in a hierarchical progression. In reality, a seven-year-old with a reading age of 7.00 will be very different, as a reader, to an eleven-year-old with a reading age of 7.00. It is wrong to believe that the reading age scale is developmental. It is also wrong to speak of reading ages like we do chronological ages. Chronological age changes at a continuous rate. Reading age does not. A reading age is specific to a subject's performance on a given date. It is misleading to describe that subject as having that reading age months later.

To provide reading ages or other derived scores, reading tests are 'normed', i.e. the test is administered to representative groups of children of different ages or different class levels (the 'norm group'). The average score for each norm group is calculated and this becomes the score that is expected or normal for a child of a particular group. However, not all children in that norm group received that score. Their scores are plotted on a normal distribution curve. While many of the children receive scores near the average, some are above and some are below that average. The test giver must understand that, as with any other characteristic, there is a natural spread of scores in any group. It is normal then that some children will score close to the average, but it is also normal that some will score below the average and some will score below the average.

To make sense of test scores, we have to know not just the average score for each group, but how wide a range of scores is normal for a particular group. We can then decide if a pupil's score is so far away from the average for his/her class or age that we need to be concerned about it. Fortunately most reading tests enable us to do this by providing tables which convert raw scores into 'standardised scores'.

Standardised scores (SS) have some important and very useful properties. The average SS is always 100. They also include a measure of spread (known as the Standard Deviation [SD]) which enables us to tell what percentages of pupils obtain scores above or below 100. Most tests have a SD of 15 points so if we know the pupil's SS and the SD of the test we can work out from the tables provided in the test manual what percentage of pupils of the same age would fall above or below the score obtained by the pupil. The table below indicates the distribution of scores of pupils of that age taking the test.

TABLE 1	
Between +1 SD and -SD (SS 115 - 85)	68% of all scores
Between + 2SD and -2 SD (SS 130 - 70)	96% of all scores
Between 0 and +1 SD (Between 0 and -1 SD)	34.13% of all scores
Between +1 and + 2SD (Between -1 and -2 SD)	13.59% of all scores
Below - 2 SD (Above + 2 SD)	2.27% of all scores

Working from the tables in the test manual, it is thus possible to translate a raw score (i.e. the number of test items answered correctly) into standard scores and the standard scores into percentiles (or centiles). By reading the appropriate table, we can find for any SS the equivalent percentile. For example, a pupil who receives a standard score of 85 would be at the 16th percentile, i.e. 16% of pupils of his/her age would score below this level; 84% would score above this level. Similarly, a pupil who receives a standard score of 115 would be at the 84th percentile, i.e. 84% of pupils of this age would score below this level, 16% would score above this level.

On a particular reading test, a nine-year-old boy might get a raw score of 48 which converts into a reading age of seven years and six months. However, according to the tables in the manual, his standardised score is at the 14th percentile. This means his performance on this test was as good as or better than 14% of children of his age. This puts his ability into perspective. While this boy may not be reading as well as the school would like him to, he would have to have a reading age of six years six months to be in the bottom 5%.

When understood, percentile scores are less liable to misinterpretation. The average range for percentile scores lies between 25 and 75. However, at the extremes minor differences between scores will be more statistically significant because of the smaller numbers of the population they relate to. For example, while there is no significant difference between a score of 50 and a score of 55, i.e. they are both definitely within the average range. The same difference of 5 between, say, 2 and 7 is very significant. This is an important statistical point which must be remembered.

Understanding the problems associated with reading ages and using standardised scores and percentiles instead can help the test giver to better identify which children really have serious problems and improve the ability to monitor their progress over time.

It is important to appreciate that however carefully tests are constructed there will be an element of error in the results they produce. All reputable tests will quote details of measurement errors [often referred to as the “standard error of measurement” (SEM)]. This allows, for any score, a confidence band to be established which indicates the probability that the pupil’s true score lies within that band. For example, if a pupil obtains a standard score of 88, and the standard error for the test is 5, we add and subtract one SEM giving a range of 83 – 93. There is a 68% chance that the pupil’s “true score” will lie within this range. If we add and subtract 10, i.e. 2 SEM, there is a 95% chance that the pupil’s true score lies within the range 78 – 98. For this reason it is good practice to report the obtained standard score (88) and follow this with the 95% confidence range, i.e. 78 – 98 in our example.

Reading Test Scales:

Note: the raw score is the actual number of test items the subject answered correctly before conversion to one of these scales.

1. Reading Ages (RA)
2. Standardised Scores

e.g.

SS	Proportion estimated to perform below SS
130	98.1
125	95.2
120	90.9
115	84.1
100	50.0
85	15.8
80	9.1
75	4.8
70	2.3

3. Percentiles

e.g. percentile percentage of pupils

performing above the score

5 th	95%	} Average range
25 th	75%	
50 th	50%	
75 th	25%	
95 th	5%	

4. Stanines: based on a scale of 9; average = 5 (the word stanine is derived from the words 'standard' and 'nine')

Source URL: *<http://www.sess.ie/node/851>*