In this part of Module Three we will explore a number of different standard scores systems. These are like percentiles in some ways - they provide a means by which test scores can be interpreted through comparison with a norm group. However, there are also some differences. In order to introduce this subject, it is first useful to look at the limitations of rank order scoring systems such as percentiles.

**PROBLEMS WITH PERCENTILES**

Although percentiles offer a user-friendly means by which test scores can be interpreted, there are some difficulties in their use, especially when making comparisons between candidates.

The problem is that small differences in raw scores are **exaggerated** when converted to percentiles if they are close to the mean for the norm group.

Differences in raw scores at the high and low extremes are **collapsed** when converted to percentiles.

In order to understand this we need to consider the shape of the normal distribution curve.

Consider four raw scores **A (20)**, **B (23)**, **C (27)** and **D (30)**. These are located as shown above in relation to the normal distribution of the norm group.

- **A** is just below the mean at the 40th percentile.
- **B** is above the mean at the 75th percentile.

However, both of the scores (A and B) are fairly close to the mean. Although the difference between them is only 3 raw scores (ie 23 - 20), when converted to percentiles there is a difference of 35 percentiles (75 - 40). This is because **most** candidates in the norm group score close to the mean rather than at the extremes.
Increasing a score from just below to just above the mean results in “jumping ahead” of many candidates who have scored close to the mean. Thus, in a rank order scoring system, small differences in raw scores close to the mean are exaggerated.

C and D are also separated by 3 raw scores (30 - 27). However, these are both located towards the top extreme of the distribution. Since very few people in the norm group score at the extremes an increase in raw scores results in “jumping ahead” of very few candidates. Hence a small difference in percentiles (95th to 99th).

A mathematical way of expressing this problem is to say that there is a non-linear relationship between raw scores and percentiles. This can be shown by plotting raw scores against percentiles on a graph:

This non-linear relationship means that it is not correct to perform mathematical calculations based on percentiles. For example, if a candidate had taken a number of aptitude tests and we wanted a measure of his/her average performance across the tests, it would not be correct to do this by averaging his/her percentile scores.

The limitations of percentiles are therefore that they can potentially distort interpretation by exaggerating differences around the mean and collapsing differences at either the low or the high extremes. Assessors using percentiles should therefore be aware of this problem. Secondly, the non-linear relationship between raw scores and percentiles means that mathematical manipulations of percentiles (such as averages) are not permissible.