Getting our heads around percentiles

Charles Darr

Is there a lot of difference between students who achieve at the 55th percentile on a test and those who achieve at the 45th? The answer might surprise you. On a Progressive Achievement Test of Reading (PAT Reading) this difference is equivalent to about two marks and could be explained as a chance occurrence, rather than a real difference in performance by two students.

Percentiles are a common and useful way to interpret a test result. However, they are often misunderstood. It is important to have a strong understanding of what they mean before using them as part of our decision making about teaching and learning.

What is a percentile?

In education, percentiles are often used to show what proportion of a particular group scored less than each possible raw score on a test. For instance, when a raw score of 27 is converted to a year group percentile of 65, it means that 65 percent of a nationally representative sample of students in that year group scored lower than 27. The most important thing to note here is that a percentile is *not* a conversion of a raw score to a score out of 100. A percentile indicates the proportion of a group who scored less than a particular raw mark.

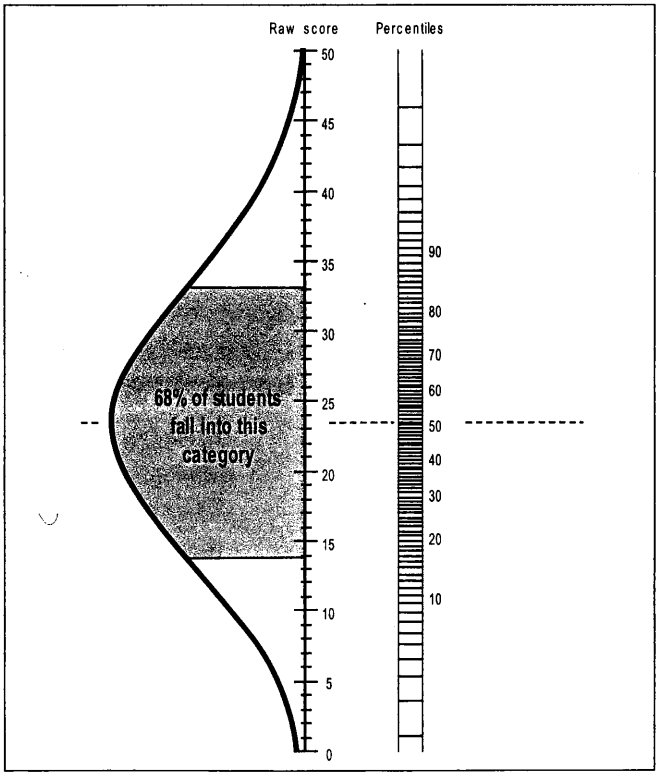
Percentiles always have to be understood in relation to a particular group

Percentiles are defined by the particular group they refer to (often called a reference group). This group is usually a particular year level or a particular age level. When using percentiles it is important to realise which group is being referred to. A student who achieves at the 75th percentile for students in Year 4 is not doing as well as a student who achieves at the 75th percentile for students in Year 5.

Percentiles do not measure achievement

Percentiles rank achievement rather than measure it. When we measure something we need to use a constant unit of measurement. Each percentile, however, does not represent the same amount of change in performance on a test. This is because more people in the reference groups score around the average mark rather than at the extremes. Figure 1 shows how the greater number of students scoring close to the average results in the percentile ranks being "crammed" together in the middle of the distribution of raw scores.

FIGURE 1 **PERCENTILES COMPARED WITH THE DISTRIBUTION OF RAW SCORES**



Percentiles can appear to indicate more difference than there really is

Because a large proportion of people score in the middle of the raw score range, an improvement of one or two marks near the test average can result in a student "leapfrogging" a large number of other students and making a marked increase in his or her percentile rank. For instance, in PAT Reading, improving by two marks somewhere near the mean results in a change of about eight percentile points. In contrast, a student who achieves very highly or very poorly on the test needs to increase his or her score by four or five marks to show the same percentile difference. A one or two mark increase will do little to change this student's ranking, as few people score at the extremes. The way percentiles bunch together in the middle explains why the percentile ranks of 55 and 45, which we began with at the start of this article, do not indicate a large difference in performance on PAT Reading.

Percentiles can give us the illusion of precision

Percentiles can sometimes make us over confident regarding the precision of a test result. All test scores are subject to measurement error. Even when we are reasonably sure that a true test result lies within a fairly tight range, for example, plus or minus three marks, this can convert to a large percentile range. For instance, using the test shown in Figure 1, a raw score of 23 with an error of plus or minus 3 indicates the true score probably lies somewhere between the 40th and 60th percentile.

Percentiles cannot be averaged

Calculating the average percentile does not produce a meaningful statistic. Averages rely on measures with a constant unit and, as has been noted, percentiles are not measures. The example below shows the misleading result of creating an average with percentiles. Here the average raw mark for a class of three students is 30. As can be seen, a raw mark of 30 is equivalent to a Year 4 percentile of 77. When we calculate the "average" percentile, however, we get 74.

	Raw score	Year 4 percentile
	35	87
	30	77
	25	57
Mean	30	74

The median percentile values *can* be used to make comparisons between two groups.

Scale scores and stanines

Percentiles provide a useful reference when comparing the achievement of students with particular reference groups. However, when we treat percentiles as a measure rather than a ranking of performance it is easy to come to the wrong conclusions about what a result means. When we come down to it, raw scores themselves should not be treated as measures. As well as, or as an alternative to percentiles and raw scores, some assessment tools report scale scores and stanines. These can be treated as measures and allow us to gauge how different achievements compare.

Charles Darr is a senior researcher and manager of the assessment design and reporting team at the New Zealand Council for Educational Research.
Email: charles.darr@nzcer.org.nz