Graphs and tables in NMSSA reporting

A reader's guide

NMSSA
Wānangatia te Putanga Tauira
National Monitoring Study of Student Achievement

This document describes how graphs are used to present findings in the reports and data windows produced by NMSSA. It also includes an explanation of some of the statistics used in tables.

Box Plots

Box and whisker plots (box plots) are used extensively throughout NMSSA reporting to summarise score distributions.

To construct a box plot, scores are ordered from low to high and then divided into four groups of equal size, called quartile groups. These are shown in Figure 1.

The box is used to show the range of the middle 50 percent of the scores and the whiskers the top and bottom 25 percent of scores. In this report, the whiskers of the box plot do not include outliers (scores considered to be unusually high or low) and have a maximum length of 1.5 multiplied by the inter-quartile (middle 50 percent) range.

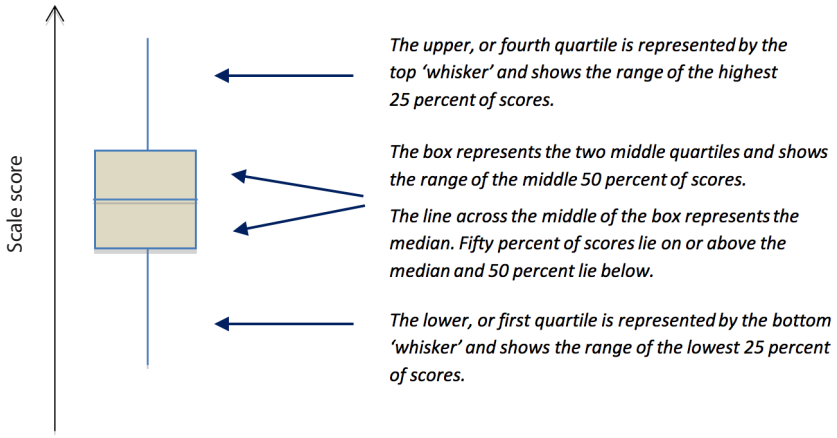

Figure 1 Understanding box plots

When box plots for two or more groups are presented as part of the same graphic, the widths of the boxes are used to represent the relative sizes of the groups. For instance, a narrow box indicates that the group size is smaller than that represented by a wider box in the same plot.

The colours for the box plots have been chosen to assist with readability. Different hues have been selected to represent each of the reported variables (for instance, gender) and two different shades of each hue chosen to represent the group at each year level (a lighter shade for Year 4 and a darker shade for Year 8).

For achievement scales that have been aligned with curriculum levels, the minimum scale score (cut-score) associated with achieving the curriculum objectives at each of the curriculum levels are indicated by horizontal dotted lines across the graph (Figure 2).

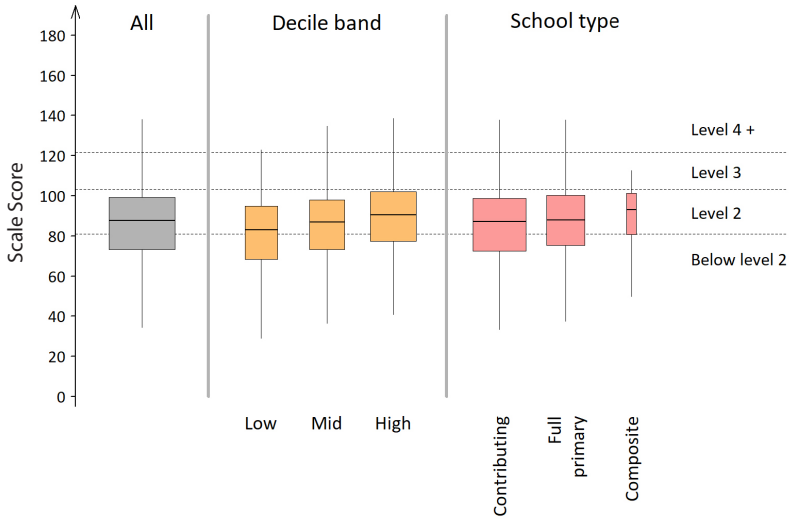

Figure 2 Interpreting box plots and NZC band levels

Line graphs of score distributions

Another type of graph used in NMSSA reporting is the line graph (Figure 3). Line graphs are used to show how the distributions of scores for Year 4 and Year 8 compare with curriculum expectations. As for the box plots, the horizontal shaded lines indicate the minimum scale scores (cut-scores) associated with achieving the curriculum objectives at each of curriculum levels 2 to 4. The shading around the lines provides a reminder that these lines represent the result of a judgement exercise (the curriculum alignment process).

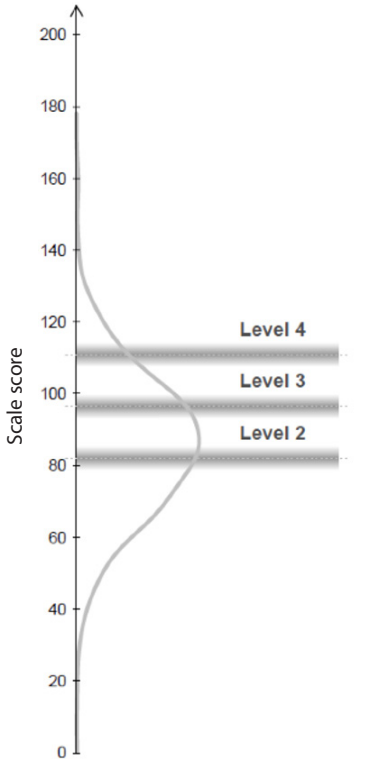

Figure 3 An example of a line graph

Tables of numerical results

Because NMSSA uses the same scale to measure achievement at both Year 4 and Year 8 it is possible to estimate how much change on average occurs on an annual basis. Table 1 gives an example of how differences in average scale scores between Year 4 and Year 8 are used to describe annualised change.

Table 1 Average difference in scale score units

	Scale score units
Difference in average scale score (Year 8–Year 4)	**29**
Confidence interval	(27.5, 30.5)
Average annual change	7.3
Average annual effect size	0.36

Table 1 also shows the 95 percent confidence interval associated with the difference in average scores at Year 4 and Year 8. Confidence intervals are used throughout reporting and provide a range within which we can be fairly sure the population value for the reported statistic lies. The confidence intervals have been adjusted to account for any design effect created through the sampling procedure (i.e., sampling schools and then sampling students). As a general rule of thumb, when the confidence intervals for two groups overlap, any difference between the groups may reasonably be explained by the kind of random variation that occurs in sampling studies (i.e., the difference between the groups is not considered to be statistically significant).

In some cases, the difference in average scores between two groups has been calculated and a confidence interval for that difference presented. When a confidence interval for a difference does not include zero, this difference can be considered to be statistically significant.

Where statistically significant differences appear in reports, they are usually bolded. For instance, in the table above, the Year 8–Year 4 difference of 29 scale score units is bolded – the difference is considered to be statistically significant.

Effect sizes are also used throughout reporting to help interpret differences between groups. An effect size quantifies the difference between the average scores for two groups in terms of standard deviation units[1]. The calculation of the effect sizes in NMSSA reporting weights the standard deviation for each group by its sample size. Because the standard deviations vary from group to group, this can mean that the same difference in scale scores can be associated with a different effect size for one pair of groups compared with another. When comparing two effect sizes it is very important to refer back to the scale score differences to make sure any interpretations are valid.

[1] *The formula for the effect size calculation is:*

$$\frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}},$$

where M1 and M2 represent the average scores for group 1 and group 2; s1 and s2 their standard deviations; and n1and n2 the number in each group.