What you need to know about data

A guide for leaders in everyday language, with templates for calculations

Data can seem scary, especially when people use terms and symbols that are unfamiliar. This guide will help you understand data using everyday language, then provide templates and clear instructions that do the calculations for you.

The big rocks – what is it testing, what does it show and is that reasonable?

The three questions below summarise just about everything that we want to know from our data. We will be using these to guide our thinking.

- *What is it testing?*
  Is this testing retention of content, transfer of learning to a new situation, solving complex problems, or something else? Have I chosen the right test for what I want to know?
- *What does it show?*
  What patterns can I find? What tends to look the same each time? What should it look like? What looks different? What doesn't fit the pattern?
- *Is that reasonable?*
  Do we have enough students for the data to be meaningful? Did we use the same conditions for both tests? Does the pattern fit with my expectations? Are we remembering that all data has an element of randomness?

Let's take each question in turn and see how it shapes our thinking.

What is it testing?

Most teachers do not realise that PAT M and NAPLAN test very different parts of maths. This means that the results from both will not necessarily align.

PAT M mostly checks if students can *remember* maths facts and procedures that they have been taught and *use* these to solve **routine questions**.
The questions are often written in a manner that is familiar to students.
It is a good test of whether students have *retained* what they have been taught over the year. It is useful as a broad measure of growth across a class, year level and region, but can also help you to identify problem areas in terms of *maths content* (e.g. lots of our students have failed in the measurement questions).

NAPLAN mostly checks whether students *understand* the important concepts in maths and whether they can *transfer or apply* this understanding to solve **challenging and complex problems**.
While a few questions are routine, most are presented in a way that is *unfamiliar* to students. Multiple choice questions often contain misconceptions in the possible answers, allowing teachers to spot underlying problems.

School-based assessment and reporting should incorporate all the proficiency strands in maths. This means that we should be able to find a pattern between our report card grades and NAPLAN results, and a pattern between our report card grades and out PAT M results, but not necessarily a pattern between our NAPLAN and our PAT M results. Examples are given in the next section.

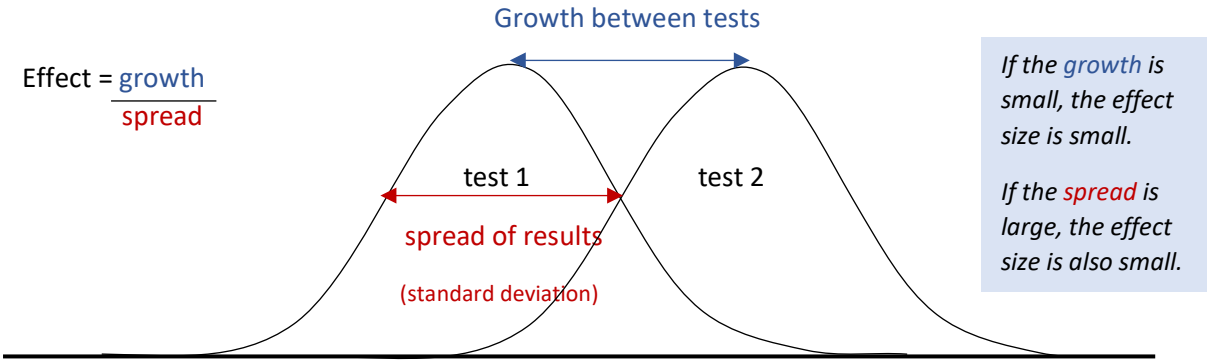
What does it show?

Once you know what is being tested, it is time to look for patterns. Both *effect size* and *graphs* are good for this. We are going to have a brief look at these below so that you know **what to look for** when you use the templates.

Effect size

Effect size is the growth in results, divided by how spread out they are. It is great for PAT M.

While the template provided will calculate effect size automatically, you need to know **what factors can mess with the results**.



Factors to consider

- If your students show a **huge range of abilities** or you have a **multi-age class**, then you are unlikely to get a big effect size. One way around this is to remove both the top and bottom student from your data set, but leave everyone else. Another option is to calculate the effect using sections of the data (e.g. split the class in half by their result on the first test, then calculate the effect for both groups separately). Effect size should apply to one year-level of students over one full year of teaching.
- If you **don't have many kids** (15 or more), your spread will be too heavily influenced by a few kids, making your effect size unreliable. In this case, you might want to combine a few classes or schools to get an accurate standard deviation, then substitute this number for your own spread in the template. Instructions for this are provided.
- You should never calculate **effect size for a single student** from only one type of testing (e.g their effect size on PAT M). There is too much potential for random error for it to be reliable. Any program that claims to do this is using data inappropriately and is not trust worthy.
- You can calculate effect size for students with low starting scores, with high starting scores, by gender, ethnicity or any other factors you like to check that you are meeting the needs of everyone. It is a reliable "overall check" to make sure that your approach is working.
- You can calculate effect size for each teacher to identify relative strengths and weaknesses for each teacher. The templates will work for every standardized test (spelling, reading etc.)

What the numbers mean

The table below should help guide your thinking.

Effect size	Negative number	-0.1 to 0.1	0.15-0.35	0.4-0.65	0.7 or higher
What this means	Students know less now than they did before.	Students have not changed since test 1.	Students have had an average year. They have not progressed by a whole year.	Students have had a good year and have improved well.	Students have improved by multiple years in a single year.

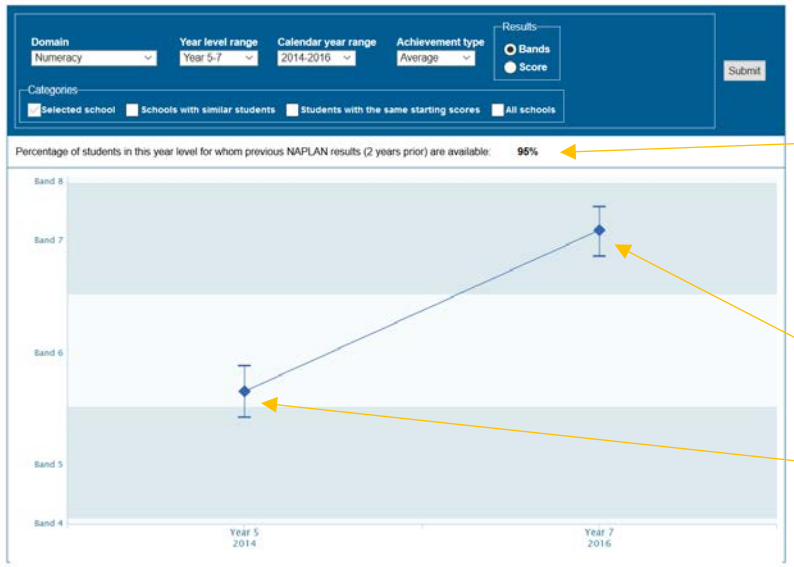
Graphing your results

Graphing is a simple way of finding patterns without having to know a lot about the numbers. If we are able to graph our results with "expected" or "normal" results then we can tell a lot more about what we are doing well and where we need to focus some attention. Below are three useful graphs that can either be found online with no effort or can be made very easily using your existing data.

MySchool student gain data

The MySchool website has great tools for looking at patterns, particularly because it can show the **growth** made by the kids you had for two years of NAPLAN testing.

- Look up your school, select "NAPLAN" from the menu on the left, then "Student gain".
- Select "Numeracy" as the domain and choose a cohort level. Submit.

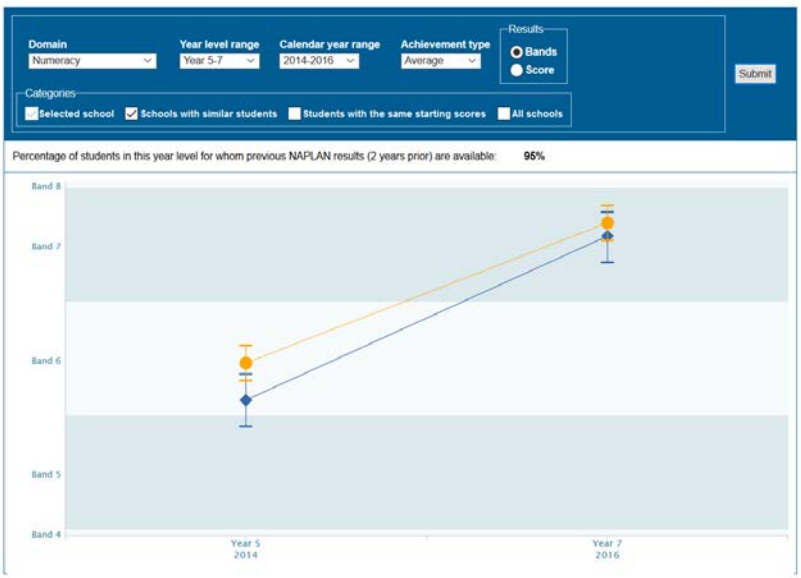


Students who sat both tests (graph doesn't show any students who were new or who left part way through)

Where they ended up in year 7

Where they started in year 5

Next, tick the box for "Schools with similar students" to compare your growth with similar schools. A yellow line will appear once you click "Submit". In the example below, the students in this school have progressed at a faster rate.

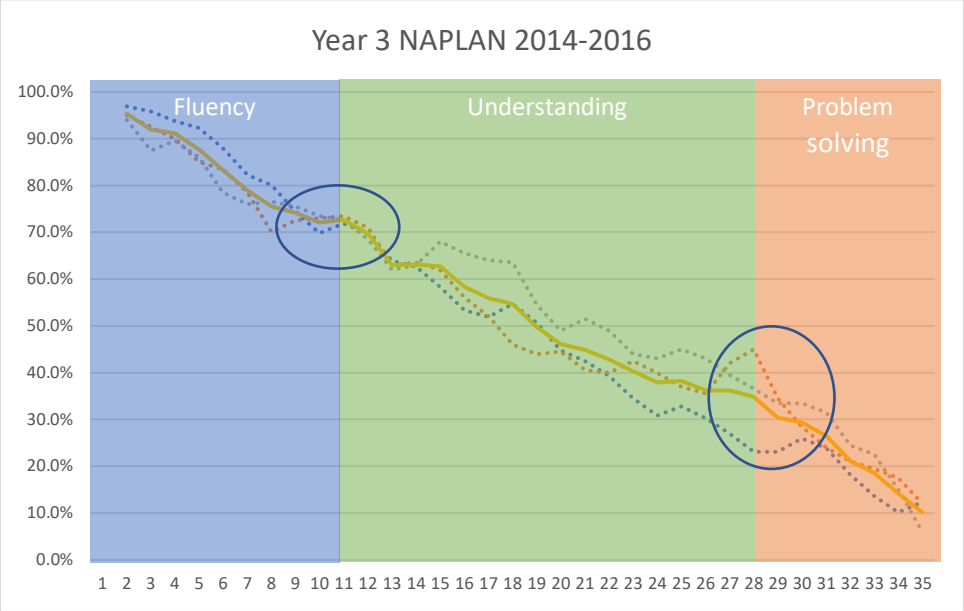


Now look for patterns:

Did this also happen in the previous years? Is it similar for reading or for writing? How does the Year 3-5 data compare to this? What doesn't fit the patterns? Why might this be the case?

NAPLAN shape graph

Below are graphs that show the normal shape of NAPLAN graphs. Each of the different years is shown by a dotted line, with the average for all three years shown by the solid yellow line. The graph shapes are consistent because of the way the tests are constructed.

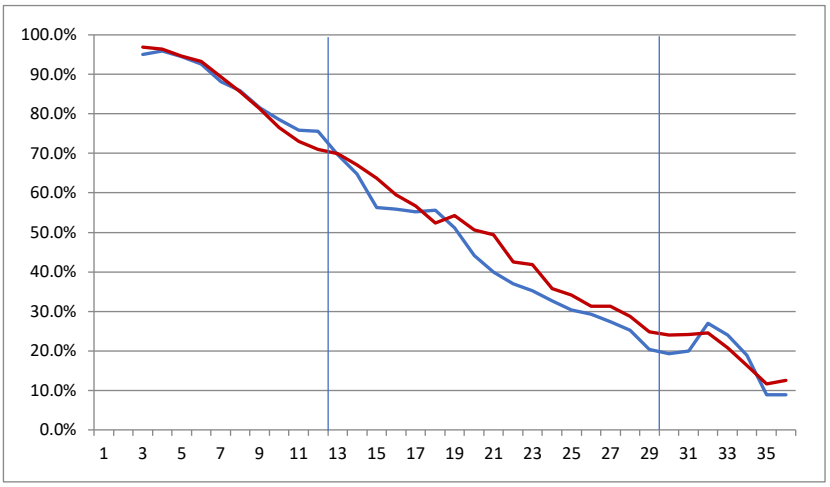


What to look for

- From question 9-13 the graphs tend to level off at around 70%. This marks the boundary between the Fluency and Understanding parts of the test.
- Another leveling or rise before a sharp fall tends to happen at roughly 30%, about 6-8 questions out from the end. This indicates where the Problem Solving questions tend to start. These questions often ask for a written response, rather than being multiple choice.
- This pattern is similar for all year levels: the first part has more Fluency, the middle part has more Understanding and the last part has more Problem Solving. Look for the leveling off followed by a sharp drop.

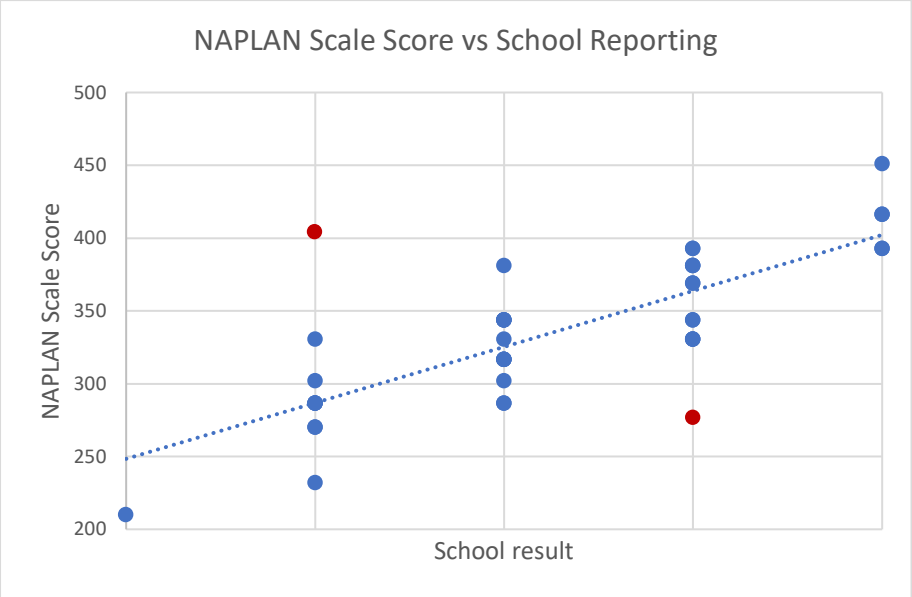
Checking your own data

Create your own data graph, with a line for your data and a line for the national data using the instructions provided. Look to see where your trendline is above, at or below the national data to determine which of the proficiencies you need to work more on. In the graph below, the school data is blue and the National data is red. What would you choose to work on? Why?



Comparing school results with NAPLAN or PAT M

It is fairly simple to create a scatter plot showing each students' scale score and their result from teachers. Adding a trendline will help you to identify any points that seem to not fit. The graph below has two points in red that do not seem to fit the trend. It is important to look at those students' results closely. What do you think has happened?



Are the results reasonable?

Any data set has the potential for some random variation – some results are lower than expected, some are higher. It is always important to ask a few questions before either discounting the results or celebrating them. Here are a few important ones:

1. Do we have enough students for this data to be meaningful? Small data sets tend to result in much higher error margins. This means that the effect sizes can look huge. The same is true for using a single student – effect size is not designed for individuals and is unreliable from only one or two tests.
2. Were the conditions "fair"? Were both tests conducted in the same manner, or did a teacher "help" more in one of the tests? E.g., PAT M growth data for year 2s fluctuates greatly depending on how the tests were conducted in year 1, so results are often not reliable.
3. Are we comparing the same group of students to each other or do we have a change in populations? If we cannot use the same population, can we compare them to a recognised "norm"?
4. Are we discounting the low achievement as "random" but not applying the same principle to high achievement?
5. Do the patterns fit with our other data, or is there a discrepancy? Is the problem with our other data or is it with this result?
6. Is this what I expected to find? If not, try to repeat the testing process to check (e.g. have a "higher than expected" result in NAPLAN sit a different NAPLAN test to confirm the result).

Final advice

Be aware of your own biases – we all tend to think that students (or staff) who are friendly are more intelligent, better at maths, or even better at teaching maths. Its always important to check our assumptions, particularly when the data paints a different picture to what we expect.