Using PAT: Mathematics to explore the consistency of teacher judgements against the National Standards

Elliot Lawes

NZCER
Rangahau Mātauranga o Aotearoa

New Zealand Council for Educational Research
PO Box 3237
Wellington
New Zealand

www.nzcer.org.nz

ISBN 978-1-927231-68-5

Contents

Tables

Figures

The focus of this paper

Teachers are uniquely placed in our education system: their skills and judgements can bring the curriculum to life and therefore have a positive influence on the learning of their students. The National Standards are intended to comprise an assessment system that recognises this unique place of teachers, and relies heavily on the skills and judgements of those teachers. The National Standards are also intended to be consistent across the nation and over time. However, the consistency of any assessment system with human judgement at its centre takes both time and resources to establish and maintain.

This paper reports on research that asked, ‘Are the National Standards a consistent assessment system?’ Our short answer is, ‘No, not yet.’ The research explored one particular aspect of consistency and focussed on whether teacher expectations for what students need to achieve in order to meet the National Standards were consistent across year levels.

In the following sections we outline the process we used to arrive at our conclusion. Key aspects of this process were:

- we used a second measure of mathematics obtained from PAT: Mathematics[1] assessment data
- we know the distributions of student achievement for this second measure of mathematics and were able to compare them to national distributions of teacher judgments against the National Standards in Mathematics

It should be noted that this second measure of mathematics is different to teacher judgment but conceptualises mathematics in a similar way.

The argument in this paper is laid out step by step so that anyone who is interested can follow it.

[1] The PAT: Mathematics assessment is used by teachers and schools throughout New Zealand. It is psychometrically robust and is explicitly linked to *The New Zealand Curriculum* (Darr, Neill, Stephanou, & Ferral, 2009). It has been developed by, and is maintained by, the New Zealand Council for Educational Research (NZCER). It has stood the test of time and it is widely trusted.

1. Overview

As part of the National Administration Guidelines, schools with students in Years 1 to 8 are required to report their teachers' judgements of student achievement in reading, writing, and mathematics to the Ministry of Education (Ministry of Education, 2013). These judgements of achievement are made against the National Standards frameworks (for example, Ministry of Education, 2010) and are categorised as either Well Below, Below, At, or Above the national standard associated with the year level of the student being judged. The judgements are called overall teacher judgements (OTJs) and are aggregated for each school. This means that no individual child's judgements are reported to the Ministry of Education.

How consistent are these OTJs? For example:

- Do teachers from one school make OTJs in a more lenient or harsh way than teachers from another school?
- Do teachers of students at one year level make OTJs in a more lenient or harsh way than teachers of students at another year level (even within the same school)?

These kinds of questions are difficult to answer because there are few opportunities for more than one teacher to make judgments about the same student. Despite the difficulty in addressing questions about the consistency of OTJs, there is some evidence to suggest that OTJs are not very consistent (Ward & Thomas, 2015).

In this report we look at the second question, rephrased as follows:

- Are teacher expectations for what students need to achieve in order to meet the National Standards consistent across year levels?

We focus on OTJs in mathematics. We have evidence that:

- teachers and curriculum experts interpret the performance expectations of the National Standards in Mathematics in different ways
- collectively, teachers apply the National Standards in Mathematics differently to students in different year levels
- compared to curriculum experts, teachers set lower thresholds for students to be judged as performing at the standards.

So, in short, teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics are not consistent across year levels.

These statements are based on an analysis of assessment data held by the New Zealand Council for Educational Research (NZCER). The assessment in question is the PAT: Mathematics assessment, a sequence of standardised tests developed specifically for use in New Zealand schools. It is psychometrically robust and was developed to measure the kind of mathematics described in *The New Zealand Curriculum* (Darr et al., 2009). PAT: Mathematics tests cover the levels of mathematics achievement typically shown by students in Years 3 to 10. The tests are all linked to what is known as the PAT: Mathematics scale. This means that every student who sits a PAT: Mathematics test is assigned a *scale score* based on the number of answers they answered correctly and the test they sat. As students' understanding of mathematics progresses, their scale scores will progress up the PAT: Mathematics scale.

NZCER holds over 450,000 PAT: Mathematics assessment records for over 200,000 students, collected from 2008 to 2013. The group of Year 4 through to Year 8 students whose assessment records are in this data set is reasonably representative of all students in these year groups so we can use this data to tell us about mathematics achievement nationally. We can also use NZCER's PAT: Mathematics assessment records to tell us about mathematics achievement against the National Standards nationally. This is because the recent development of the Progress and Consistency Tool (PaCT) provides an explicit linear relationship between PAT: Mathematics and the performance expectations of the National Standards. It is this relationship that allows us to use the distributions of student achievement in PAT: Mathematics to compare teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics across year levels.

Our examination of NZCER's PAT: Mathematics data and the national distribution of OTJs is not complex, but it does progress in several stages. These are laid out in a step-by-step way in sections 2 through 7. Some discussion follows in section 8.

2. OTJs in mathematics for 2013

Table 1 shows the 2013 national distribution of OTJs in mathematics. This table is based on public achievement information data held by the Ministry of Education.

Table 1 **National distribution of overall teacher judgements for 2013**

Year group	Percentage in National Standards reporting category (%)			
	Well Below	Below	At	Above
Year 4	5	18	52	25
Year 5	6	22	50	22
Year 6	6	18	50	26
Year 7	9	26	43	23
Year 8	8	23	43	26

If we scan down the left-most column of numbers in Table 1, we can compare the percentages of students from different year levels being judged Well Below the standard. These percentages don't vary much. Similarly, if we scan down the centre-left column of numbers in Table 1, we can compare the percentages of students from different year levels being judged Below the standard. These percentages also don't vary much.

In fact we see this for each of the columns of numbers. Regardless of the year level of the students:

- between 5 and 9 percent of students were judged Well Below their year level's standard
- between 18 and 26 percent of students were judged Below their year level's standard
- between 43 and 52 percent of students were judged At their year level's standard
- between 22 and 26 percent of students were judged Above their year level's standard.

At this point it is difficult to tell what this year-to-year stability means. It is tempting to think that it means that teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics are consistent across year levels. In sections 6 and 7 we will see that this is not the case.

For now, in order to understand more about the percentages in Table 1 and their year-to-year stability, we need to understand more about the PAT: Mathematics assessment.

3. The PAT: Mathematics assessment

The PAT: Mathematics assessment is a sequence of standardised tests developed specifically for use in New Zealand schools. The tests are designed to help classroom teachers to understand the achievement and progress of their students in mathematics as it is described in *The New Zealand Curriculum* (Ministry of Education, 2007).

PAT: Mathematics tests cover the levels of mathematics achievement typically shown by students in Years 3 to 10. The tests are all linked to what is known as an *equal interval scale*. Every student who sits a PAT: Mathematics test is assigned a *scale score* based on the number of answers they answered correctly and the difficulty of the test they sat. The fact that the scale is an equal interval scale means that progress of one unit up the scale indicates the same amount of mathematical progress no matter where on the scale the progress occurs.

As an example, a Year 4 student who scored 50 percent correct on the Year 4 test they sat will earn a scale score of 29.7 scale units, whereas a Year 8 student, who also scored 50 percent correct on the much more difficult Year 8 test they sat, will earn a scale score of 56.3 scale units. While the percentage correct of the two students is the same, the two students are clearly at different levels of mathematics achievement (because one of the tests was much more difficult than the other). This differing amount of mathematics achievement *is* captured by the PAT: Mathematics scale score.

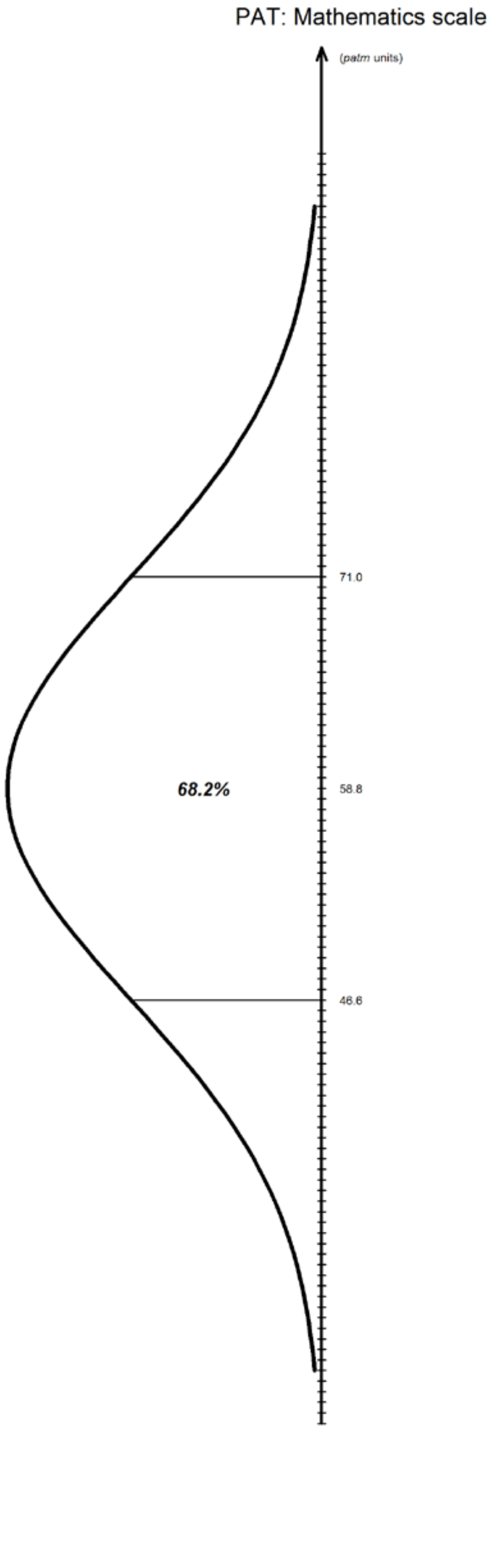
The scale is known as the PAT: Mathematics scale and its units are called *patm* units. The PAT: Mathematics scale is analogous to a temperature scale, and the *patm* units are analogous to ºC, which are units of measurement on a scale of temperature.

The PAT: Mathematics scale allows us to look at patterns in the distributions of student achievement for students in different year levels. For example, the distribution of PAT: Mathematics scale scores of Year 8 students on the PAT: Mathematics scale is shown in Figure 1.

Figure 1 plots the distribution of achievement for Year 8 students on the PAT: Mathematics scale using results from NZCER's assessment data for tests completed in Term 4. The distribution is normal (i.e. bell-shaped), with an average of 58.8 *patm* and a standard deviation (an indicator of variability) of 12.2 *patm*. We can also see that around 68.2 percent of Year 8 students scored between 71.0 patm (one standard deviation above the mean) and 46.6 *patm* (one standard deviation below the mean). These numbers are slightly different to the Term 1 distributions in Darr et al., 2009: we use Term 4 records from NZCER's assessment data because it is in Term 4 that OTJs are made.

But what does this tell us about the percentages in Table 1? What does it tell us about the year-to-year consistency of teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics?

Figure 1 **The distribution of Year 8 students' Term 4 PAT: Mathematics scores**



4. Converting OTJs to the PAT: Mathematics scale

This section returns to the issue of trying to understand the percentages in Table 1 and what they tell us about the year-to-year consistency of teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics. This question can be addressed by relating the percentages in Table 1 to distributions of student mathematics achievement measured by the PAT: Mathematics assessment.

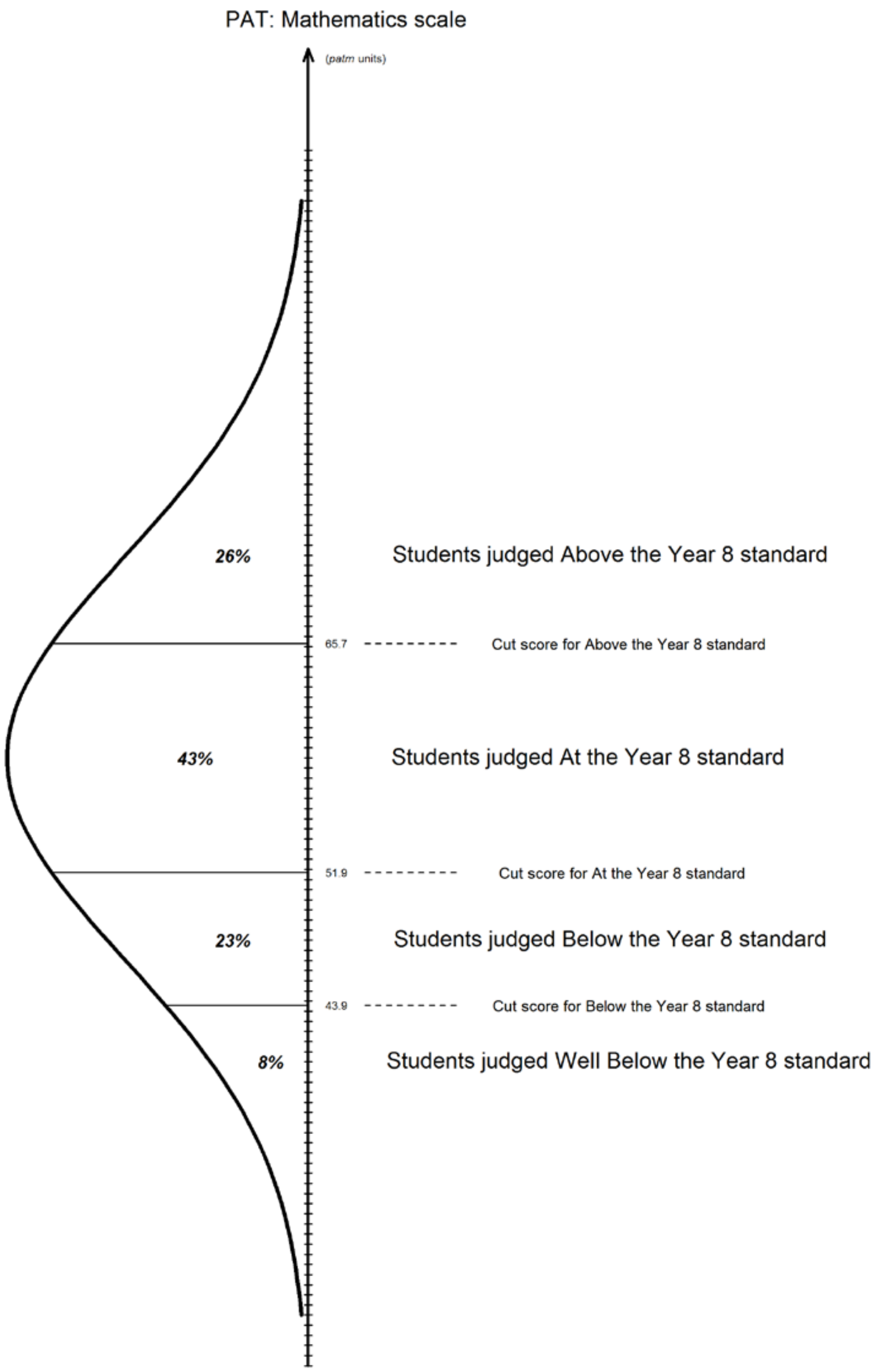
For example, let's look at the last row of Table 1. This is the row describing the mathematics OTJs given to Year 8 students in 2013. It shows that:

- the mathematics performance of around 8 percent of Year 8 students was judged Well Below the Year 8 standard
- the mathematics performance of around 23 percent of Year 8 students was judged Below the Year 8 standard
- the mathematics performance of around 43 percent of Year 8 students was judged At the Year 8 standard
- the mathematics performance of around 26 percent of Year 8 students was judged Above the Year 8 standard.

This is one way of describing the distribution of mathematics achievement of Year 8 students. As we know from Figure 1, NZCER's PAT: Mathematics assessment data give us another way.

We know that the mathematics achievement of around 8 percent of Year 8 students was judged Well Below the Year 8 standard. We can show this percentage on the distribution of PAT: Mathematics scores for Year 8 students. Similarly, we know the percentages of Year 8 students judged Below, At, and Above the Year 8 standard. We can also show these percentages on the distribution of PAT: Mathematics scores for Year 8 students. These percentages are shown in Figure 2.

Figure 2 **The distribution of Year 8 students' PAT: Mathematics scores together with cut scores for the Year 8 standard**



Also shown in Figure 2 are the scores on the PAT: Mathematics scale that separate students judged at one reporting band from students judged at another. These scores are known as *cut scores*.

We can think of cut scores as the scores on the PAT: Mathematics scale that represent the minimum performance necessary to be judged at a particular National Standards reporting category. In Figure 2 it is the teachers of Year 8 students collectively who are in effect ‘setting the cut scores’.

In Figure 2 we see that teachers of Year 8 students have collectively determined that:

- the cut score representing the minimum performance necessary to be judged Below the Year 8 National Standard in Mathematics is 43.9 *patm* units
- the cut score representing the minimum performance necessary to be judged At the Year 8 National Standard in Mathematics is 51.9 *patm* units
- the cut score representing the minimum performance necessary to be judged Above the Year 8 National Standard in Mathematics is 65.7 *patm* units.

Notice that there is one fewer cut score than there are reporting categories. This is because Well Below is the lowest reporting category and there is no minimum performance necessary to be judged Well Below the Year 8 standard in mathematics.

So we have transformed the distribution of OTJs given by teachers of Year 8 students into three cut scores on the PAT: Mathematics scale. But there was nothing special about Year 8: we can do this for all of the year levels for which we have data. Table 2 shows what happens if we transform the distributions of OTJs given by teachers of students from all of the Year levels specified in Table 1.

Table 2 **PAT: Mathematics cut scores determined by the national distribution of overall teacher judgements for 2013**

Year-level standard	**Reporting band cut score (*patm* units)**		
	Below	**At**	**Above**
Year 4	18.9	28.9	46.2
Year 5	28.1	37.3	53.3
Year 6	33.1	40.7	56.5
Year 7	40.1	48.8	62.6
Year 8	43.9	51.9	65.7

We can see that the bottom row of Table 2 shows the Year 8 cut scores that we first saw in Figure 2.

But what does Table 2 tell us about the percentages in Table 1 and their year-to-year stability? To answer this question we still need to understand how the reporting bands for the standard at one year level relate to the reporting bands for the standard at another year level. We will discuss this in section 5. However, for now it is worth noticing that the year-to-year stability that we saw in the percentages in Table 1 manifests in Table 2 as year-to-year stability in the differences between cut scores.

Specifically, if we calculate the difference between the At cut score for the Year 4 standard (28.9 *patm* units) and the Below cut score for the Year 4 standard (18.9 *patm* units), we get 10 *patm* units. If we repeat this calculation for the Year 5 standard we get a difference of 9.2 *patm* units. This difference does not vary much over all of the Year levels shown in Table 2.

Similarly, if we calculate the difference between the Above cut score for the Year 4 standard (46.2 *patm* units) and the At cut score for the Year 4 standard (28.9 *patm* units), we get 17.3 *patm* units. Again, if we repeat this calculation for all of the Year levels shown in Table 2, the resulting difference does not vary much. These calculations are shown in the left-hand and central columns of numbers in Table 3.

Table 3 **Differences between cut scores**

Year-level standard	Difference between At and Below cut scores (*patm* units)	Difference between Above and At cut scores (*patm* units)	Ratio of differences
Year 4	10.0	17.3	1.7
Year 5	9.2	16.0	1.7
Year 6	7.6	15.8	2.1
Year 7	8.7	13.8	1.6
Year 8	8.0	13.8	1.7

The right-hand column in Table 3 shows that the differences between the Above and At cut scores are consistently between 1.6 and 2 times as large as the differences between the At and Below cut scores. In section 7 we will see that this fact has important consequences. But to fully understand this, we first need to understand how the reporting bands for the standard at one year level relate to the reporting bands for the standard at another year level. This will allow us to return to Table 1 and what it says about the consistency of teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics.

5. The relationship between OTJs for different year levels

So how do the reporting bands for the standard at one year level relate to the reporting bands for the standard at another year level? For example, if the mathematics achievement of a Year 4 student was judged At the Year 4 standard, how should it be judged against the Year 5 standard? This is an important question because teachers are asked to think across year levels when making an OTJ.

According to a Ministry of Education's fact sheet:

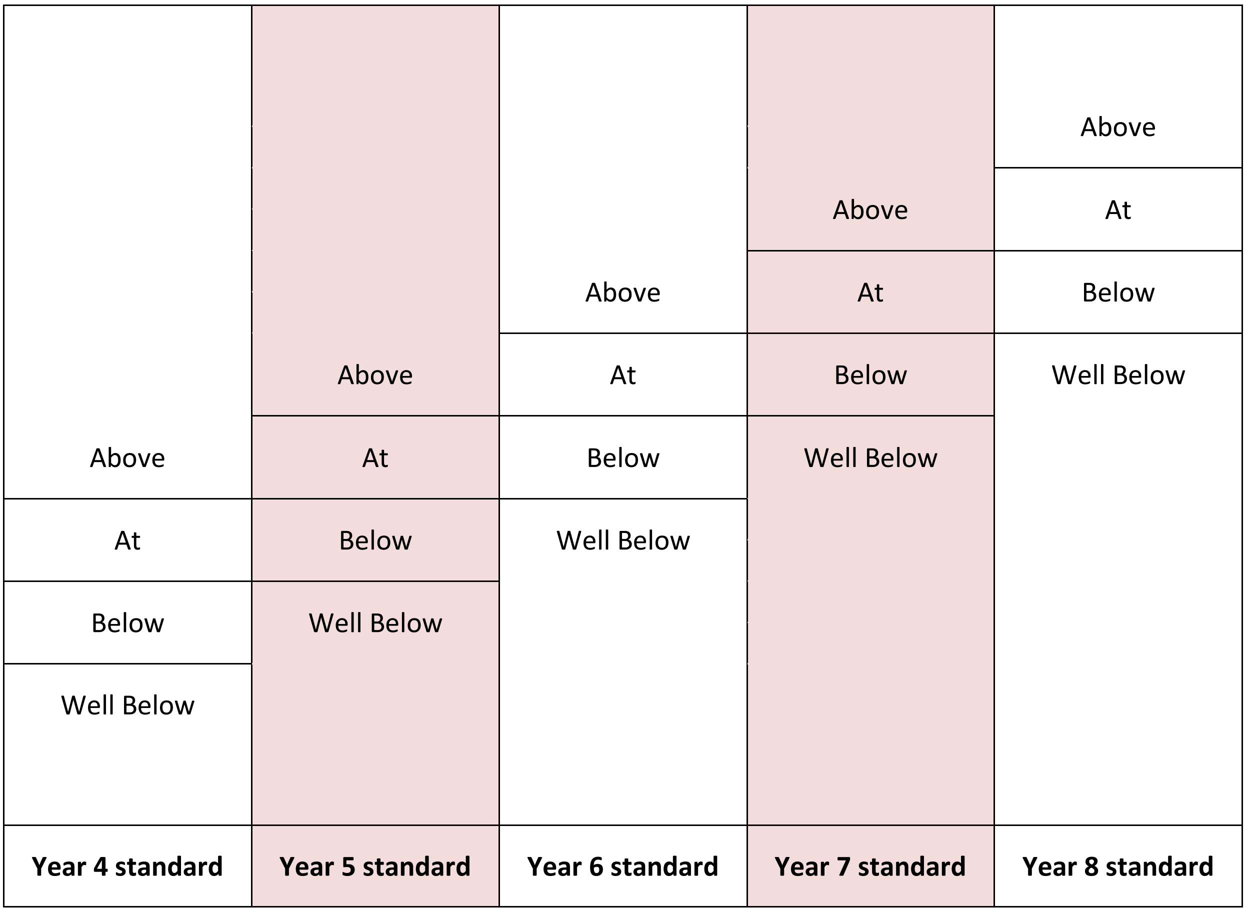
> If the balance of evidence shows the student's achievement is:
>
> - in a year level above a National Standard, the student's achievement will be described as above the National Standard
> - predominantly meeting the expectations at a year level, the student's achievement will be described as at the National Standard
> - not achieving at a National Standard, but achieving closer to the National Standard immediately below, the student's achievement will be described as below the National Standard
> - more than one year below a National Standard, the student's achievement will be described as well below the National Standard.
>
> (Ministry of Education, 2011)

What this is saying is summarised in Figure 3. The key point is that At the standard for one year level aligns with Below the standard for the following year level. What does this mean for cut scores? We can see this in Figure 3 by looking at the horizontal lines.

The horizontal line representing the Above cut score for the Year 5 standard is the line in Figure 3 that is in the column labelled 'Year 5 standard' and that sits between the words 'Above' and 'At'. We can see that it extends across the columns representing the Year 6 and Year 7 standards. In the column representing the Year 6 standard it sits between the words 'At' and 'Below'—here it represents the At cut score for the Year 6 standard. In the column representing the Year 7 standard it sits between the words 'Below' and 'Well Below'—here it represents the Below cut score for the Year 6 standard. So we can see that the Above cut score for the Year 5 standard should be equal to the At cut score for the Year 6 standard, which should in turn be equal to the Below cut score for the Year 7 standard. We can apply this interpretation across all of the horizontal lines in Figure 3.

Now we can return to Table 2 and use this new understanding of how the cut scores are intended to relate to one another. This will allow us to understand the percentages in Table 1 and what they say about the year-to-year consistency of teacher expectations for what students need to achieve in order to meet the National Standards in Mathematics.

Figure 3 **How the reporting bands for the standard at one year level relate to the reporting bands for the standard at another year level**



6. How OTJs for different year levels are related on the PAT: Mathematics scale

Now that we know how the reporting bands and associated cut scores for one year-level standard relate to those for another year-level standard, we can transform Table 2 to reflect this relationship.

To help understand the steps involved, we first express just the Year 5 cut scores in terms of the Year 4 cut scores. We know from Figure 3 that the cut score for being at the Year 4 standard should be equal to the cut score for being below the Year 5 standard. We also know that the cut score for being above the Year 4 standard should be equal to the cut score for being at the Year 5 standard. Table 4 shows the result of our transformation.

Table 4 **Transforming cut scores: 1**

Year group	Cut score for each year-level standard (*patm* units)			
	Below 4	At 4	Above 4	
		Below 5	At 5	Above 5
Year 4	18.9	28.9	46.2	
Year 5		28.1	37.3	53.3

Table 4 is still a little confusing because each column of cut scores has two labels: one inherited from the Year 4 standard and one inherited from the Year 5 standard. But if we decide to label each column with the standard it is at, we should get rid of this confusion. We just have to remember that:

- the cut score for below the Year 4 standard should be equal to the cut score for being at the 'after 3 years at school' standard
- the cut score for being above the Year 5 standard should be equal to the cut score for being at the Year 6 standard.

After relabelling, Table 4 looks like Table 5.

Table 5 **Transforming cut scores: 2**

Year group	Cut score for each year level standard (*patm* units)			
	At after 3 years	At 4	At 5	At 6
Year 4	18.9	28.9	46.2	
Year 5		28.1	37.3	53.3

We can continue this for the other years in Table 2. If we do this, we end up with Table 6.

Table 6 **PAT: Mathematics cut scores determined by the national distribution of overall teacher judgements for 2013**

Year group	Cut scores for each year-level standard (*patm* units)						
	At after 3 years	At 4	At 5	At 6	At 7	At 8	Above 8
Year 4	18.9	28.9	46.2				
Year 5		28.1	37.3	53.3			
Year 6			33.1	40.7	56.5		
Year 7				40.1	48.8	62.6	
Year 8					43.9	51.9	65.7

How should we read Table 6? As an example, let's look at the row corresponding to Year 6. The three cut scores on this row are 33.1 *patm* units, 40.7 *patm* units and 56.5 *patm* units. The cut score of 33.1 sits in the column under the heading 'At 5'. This means that the teachers of Year 6 students made OTJs in such a way that they effectively set the cut score for the 'at the Year 5 standard' at 33.1 *patm* units. Similarly, the cut score of 40.7 sits in the column under the heading 'At 6'. This means that the teachers of Year 6 students made OTJs in such a way that they effectively set the cut score for the 'at the Year 6 standard' at 40.7 *patm* units. Finally, the cut score of 56.5 sits in the column under the heading 'At 7'. This means that the teachers of Year 6 students made OTJs in such a way that they effectively set the cut score for the 'at the Year 7 standard' at 56.5 *patm* units. We can read the other rows of Table 6 in a similar way.

How should we *understand* Table 6? The first thing to notice is that the cut scores increase across the rows. This is simply an artefact of how we constructed the cut scores (described in section 4). It reflects our assumption that the cut score for below the standard should have a lower value than the cut score for at the standard, which should in turn have a lower value than the cut score for above the standard.

The second thing to notice about Table 6 is that the cut scores decrease moving down the columns. This is *not* an artefact of how we constructed the cut scores. It means that for a given standard, teachers of students at higher year levels tend to set lower cut scores for that standard than teachers of students at lower year levels. In other words, teachers of students at higher year levels have lower expectations of the performance levels required to meet the National Standards in Mathematics than do the teachers of students at lower year levels.

It is worth noting that here we are describing the expectations of one set of teachers *relative* to the expectations of another set of teachers. We are not yet saying which set of expectations is better, or more closely aligned with the expectations of curriculum experts. We will discuss that in section 7.2.

Is it possible to understand this difference in cut scores in a more practical sense? One way to do this is by expressing the difference in cut scores from Table 6 in terms of how much progress in mathematics achievement a typical student makes in a year. To do this we need to know the average Term 4 achievement in *patm* units of students at different year levels. These are provided in Table 7.

Table 7 **Average achievement in Term 4, by year level in PAT: Mathematics**

Year level	Average PAT: Mathematics scale score (*patm* units)	Average progress since previous year level (*patm* units)
Year 4	38.1	NA
Year 5	44.7	6.6
Year 6	49.4	4.8
Year 7	54.6	5.2
Year 8	58.8	4.2

The numbers in the right-hand column of Table 7 are simply the differences between successive entries in the left-hand column. The numbers in the right-hand column represent the average progress that students make over a year in terms of *patm* units.

We can now use Table 7 to interpret Table 6. For example, in Table 6 we see that the teachers of Year 6 students set the cut score for the At Year 7 standard at 56.5 *patm* units. In Table 6 we also see that teachers of Year 8 students set the same cut score at 43.9 *patm* units. The difference between these is about 13 *patm* units. Looking at Table 7, we can see that this difference is equivalent to 2 or more years' progress on average up the PAT: Mathematics scale: a substantial difference in terms of meeting the requirements to be judged At a standard.

7. Conclusions

So what can we conclude from what we've shown? Recall from section 6 that teachers of students at higher year levels appear to have lower expectations for what students need to achieve in order to meet the National Standards in Mathematics than do the teachers of students at lower year levels. Recall also that for some standards, this difference in expectations was equivalent to at least 2 years' worth of growth in mathematics achievement.

This clearly answers our question 'Are the National Standards a consistent assessment system?' They are not yet. We will now explore what we have shown in a little more detail.

7.1. Collectively, teachers apply the National Standards in Mathematics differently to students from different year levels

We know that teachers of students at higher year levels have lower expectations for what students need to achieve in order to meet the National Standards in Mathematics than do teachers of students at lower year levels. It must follow that, collectively, teachers apply the National Standards in Mathematics differently to students from different year levels. But what was the key to us discovering this? Recall from section 2 that there was year-to-year stability in the percentage of students judged in each National Standards reporting category. We were tempted to think that this implied year-to-year consistency in the way the teachers apply the National Standards. But we were able to check this because we had another way of describing student achievement in the same subject.

7.2. Compared to curriculum experts, teachers set lower thresholds for students to be judged as performing at the standards

As we mentioned in section 4, across all of the year levels the differences between the Above and At cut scores were consistently between 1.6 and 2 times as large as the differences between the At and Below cut scores. Because the PAT: Mathematics scale is an equal interval scale, this difference can be interpreted as a real difference in terms of student achievement. That is, within a year level, teachers are interpreting the National Standards in such a way that if we were to compare students who had just qualified to be judged Above the Standard with students who had just qualified to be judged At the standard, we would find a substantially greater difference in mathematics achievement than if we were to compare students who had just qualified to be judged at the Standard with students who had just qualified to be judged below the standard.

By itself this is not problematic. But when students from one year are compared with students from another year, it becomes problematic. This is because the OTJ reporting bands for one year are related to those for another year (as described in section 5), and so the student achievement they describe should be related in a coherent way. In Table 6 we see that this student achievement is not related coherently. For any of the columns in Table 6 where there are three cut scores, it is always the cut score corresponding to the youngest year level (the one at the top in the column) that stands out. This is the cut score corresponding to the Above reporting category (the middle number of the three corresponds to the At reporting category and the bottom number of the three corresponds to the Below reporting category).

Because we are describing the location of cut scores *relative* to one another, either or both of the following statements must be true:

- teachers set low thresholds for students to be judged At or Below the standards
- teachers set high thresholds for students to be judged Above the standards.

To understand whether one of these statements is more appropriate than the other, we will now look at how curriculum experts would set cut scores for the National Standards in Mathematics on the PAT: Mathematics scale. If one of the above statements aligns more closely with the views of curriculum experts, then we will adopt that statement of the relative difference between cut scores.

In 2014, a group of curriculum experts in mathematics was convened to set the standards on the Ministry of Education's Progress and Consistency Tool (PaCT) (Ministry of Education, 2015). One goal of the PaCT tool is to improve the consistency of OTJs. It aims to do this by providing teachers with a richly illustrated judgement framework, together with a scoring mechanism linked to an equal interval scale and recommendations for which levels of judgement correspond to which standards. The curriculum experts in mathematics contributed directly to the issue of which levels of judgement correspond to which standards: they set cut scores associated with each of the standards. Standard-setting is an activity that can highlight many issues associated with standards (NZCER, 2010). The standard setting for PaCT is no exception.

During the development of the PaCT, an explicit link was determined with the PAT: Mathematics scale (Lawes & Darr, 2014). This allows the standards that the curriculum experts set for PaCT—in the form of cut scores—to be transferred to the PAT: Mathematics scale and compared with the standards effectively set by teachers in Table 6.

These cut scores, transformed onto the PAT: Mathematics scale, are shown in Table 8.

Table 8 **PaCT standard-setting cut scores determined by curriculum experts**

PAT: Mathematics cut score for year level standards (*patm* units)						
At after 3 years	**At 4**	**At 5**	**At 6**	**At 7**	**At 8**	**Above 8**
32.8	38.3	43.7	49.7	54.8	61.6	68.3

For comparison, Table 6 is replicated below.

Table 6 **PAT: Mathematics cut scores determined by the national distribution of overall teacher judgements for 2013**

Year group	Cut score for each year-level standard (*patm* units)						
	At after 3 years	At 4	At 5	At 6	At 7	At 8	Above 8
Year 4	18.9	28.9	46.2				
Year 5		28.1	37.3	53.3			
Year 6			33.1	40.7	56.5		
Year 7				40.1	48.8	62.6	
Year 8					43.9	51.9	65.7

When we compare these tables we see that for the standards above the At 4 standard, the cut score set by the curriculum experts more closely aligns with the cut score corresponding to the Above reporting category (the top cut score in each column) than the cut scores corresponding to the At reporting category (the middle cut score in each column) or the cut scores corresponding to the Below reporting category (the bottom cut score in each column).

Therefore, we can conclude that, compared with curriculum experts, teachers set low thresholds for students to be judged At the standards. It is equally true that compared with curriculum experts, teachers set low thresholds for students to be judged Below the standards.

7.3. Teachers and curriculum experts interpret the performance expectations of the National Standards in Mathematics in different ways

It is worth comparing Table 6 and Table 8 further. The first thing we notice when we compare them is that there is only one cut score for every standard in Table 8 (i.e. just one row of cut scores). This may seem trivial. It is a consequence of how the curriculum experts set the standards: they worked over several standards at the same time, making sure that their understanding of one standard cohered with their understanding of another. But while trivial, this difference between Table 6 and Table 8 means that:

- the understanding of the National Standards that curriculum experts had, and that their process *required* them to have, ensured a coherent set of relationships between every standard and every other standard, whereas
- the collective understanding of the National Standards that teachers had, and that the OTJ process *allowed* them to have, permitted a less-than-coherent set of relationships between every standard and every other standard.

This certainly implies that teachers and curriculum experts interpret the performance expectations of the National Standards in Mathematics in different ways.

It is worth noting that when the curriculum experts were setting the standards, they largely agreed about where the cut scores should be. But their agreement wasn't complete: the cut scores in Table 8 are an average of the curriculum experts' individual cut scores. So the process used by the curriculum experts, and the understanding of the National Standards they had, does not remove possible inconsistency between different people applying the standards. But it *does* remove possible inconsistency between applying the standards to student achievement from different year levels.

8. Discussion

It might be tempting to interpret the main findings of this paper (presented in section 1) as an indictment of the ability of teachers to collectively provide consistent judgements against the National Standards. Obviously consistency is highly desirable, but this interpretation ignores what teachers *do* contribute: deep knowledge of their students and the environments in which those students learn. While we should have high expectations for teacher consistency, we should also have realistic ones and remember that consistency is not the only desirable outcome for an assessment system. What, then, does this paper tell us?

This paper highlights the issue of the way the standards are required to be reported. Teachers of students in Years 1 to 8 are required to report their judgements of student achievement as either Well Below, Below, At or Above the national standard associated with the year level of the student being judged. But there is very little guidance provided to teachers on how these reporting bands relate to one another—especially from one standard to the next. It is difficult to guess the impact that requiring teachers to use these four labels (Above, At, Below, and Well Below) when reporting has on year-to-year consistency of OTJs. However, it is equally difficult to imagine that it has no impact at all.

This paper also demonstrates that the OTJ data held by the Ministry of Education for students in one year level (as in Table 1) has an inconsistent relationship with the OTJ data held by the Ministry of Education for students in other year levels. If judgements or decisions about the quality of any parts of New Zealand's primary education system are going to be made using this data, the data should have reasonable levels of year-to-year consistency (as well as teacher-to-teacher and school-to-school consistency). Otherwise, how will we know if the judgement or decision reflects reality or merely how poorly we are describing reality? If this OTJ data is an important indicator of New Zealand's primary education system, it is vital that we monitor its consistency. One way of doing so—as we have done in this paper—is to use another measure of student achievement (in the same subject) for which we know the distribution.

The kind of investigation that we have carried out in this paper is never going to be able to probe the cause of the inconsistency it has revealed, so we are left to speculate. Of course it is teachers who make the actual judgements against the National Standards, but they do so in an environment in which it is extremely difficult to encourage the kind of year-to-year consistency that has been the focus of this paper. In particular, many teachers in New Zealand teach students from just one or two year levels. Those teachers who are willing and able, moderate their judgments in the first instance with their colleagues who teach students of the same year level. But how many have the luxury of moderating their judgments with colleagues who teach students from a different year level?

A tool such as the PaCT does support teachers to make judgements in a way that is consistent from year to year, but the use of the PaCT is not compulsory and teachers are not required to use the OTJ that it recommends. And even if they were, there is still plenty of scope for teachers to interpret the PaCT frameworks differently (Lawes & Darr, 2014). It is more reasonable to think that the kind of year-to-year consistency this paper has focused on could emerge from the judicious use of resources (such as the PaCT tool), in combination with strong professional development for teachers that is centred on both judging and understanding student achievement in various curriculum areas. Both the use of resources and professional development for teachers will require time to allow teacher practice to evolve and grow stronger.

While the National Standards are an assessment system that has not yet reached its intended level of consistency across the nation and over time, the Standards *do* have the skills and judgements of teachers at their heart and are able to capture those teachers' knowledge of their students.

References

Darr, C., Neill, A., Stephanou, A., & Ferral, H. (2009). *Progressive Achievement Test: Mathematics, Teacher manual, 2nd Edition, Revised.* Wellington: New Zealand Council for Educational Research.

Lawes, E., & Darr, C. (2014). *The Progress and Consistency Tool: Mathematics framework technical report.* Wellington: New Zealand Council for Educational Research.

Ministry of Education. (2007). *The New Zealand curriculum: For English-medium teaching and learning in years 1–13.* Wellington: Learning Media.

Ministry of Education. (2010). *NZC mathematics standards for years 1−8 poster.* Wellington: Learning Media Ltd. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Mathematics-standards/Supporting-resources

Ministry of Education. (2011). *Overall teacher judgment.* Wellington: Author. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/Overall-teacher-judgment#4

Ministry of Education. (2013). *The National Administration Guidelines (NAGs): National administration guideline 2A.* Wellington: Author. Retrieved from http://www.education.govt.nz/ministry-of-education/legislation/nags/#NAG2A

Ministry of Education. (2015). *The Progress and Consistency Tool (PaCT): Supporting professional judgments in reading, writing, and mathematics.* Wellington: Author. Retrieved from http://pactinfo.education.govt.nz/

New Zealand Council for Educational Research. (2010). *National Standards for Mathematics: Where to set the bar?* Wellington: Author. Retrieved from http://www.nzcer.org.nz/research/publications/national-standards-mathematics-where-set-bar

Ward, J., & Thomas, G. (2015). *National Standards: School sample monitoring and evaluation project, 2010−2013.* Wellington: Ministry of Education. Retrieved from http://www.educationcounts.govt.nz/publications/series/National_Standards/national-standards-school-sample-monitoring-and-evaluation-project-2010-2013