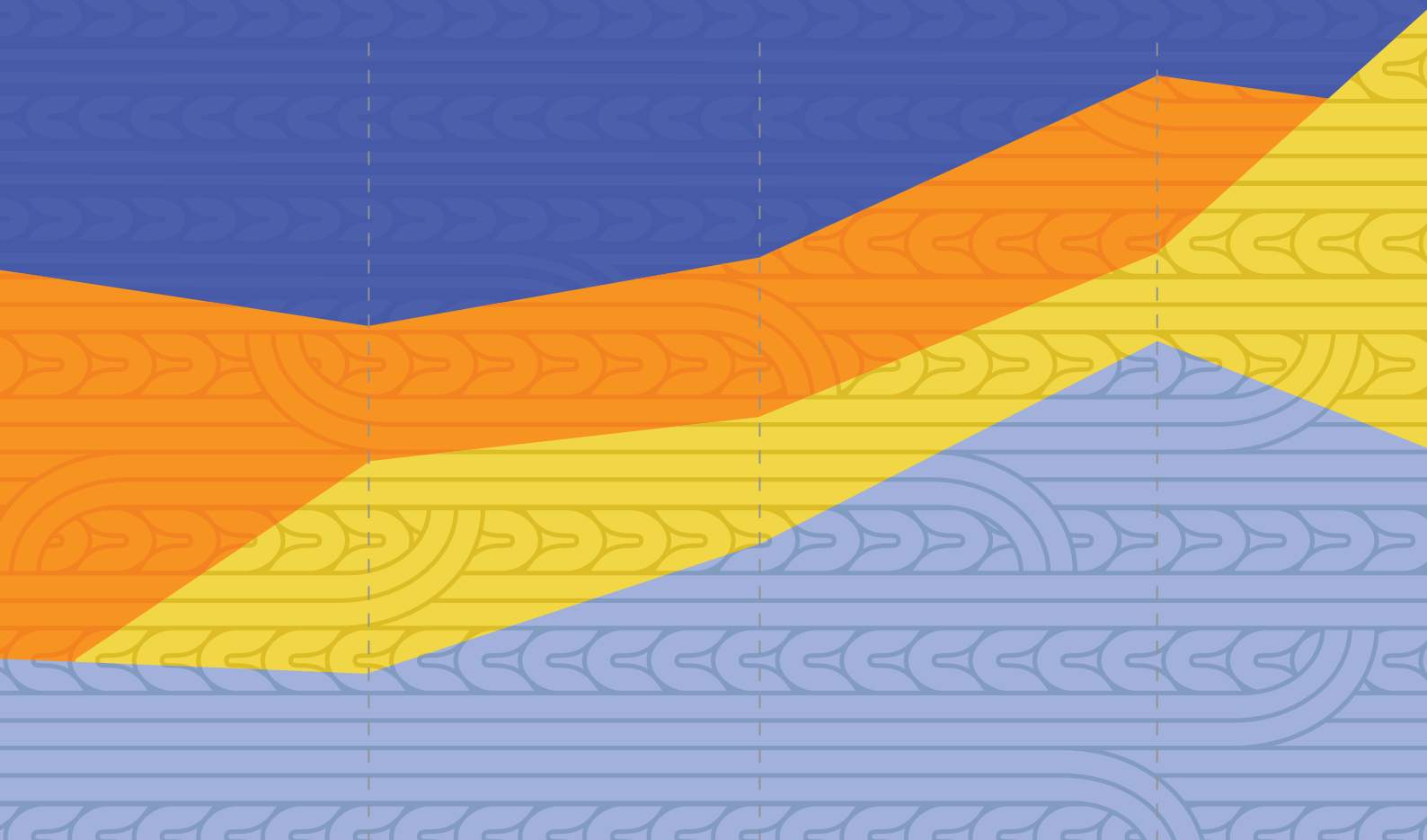


PAT trends in Aotearoa New Zealand

Evidence in mathematics,
reading comprehension,
and reading vocabulary
from 2012 to 2026

David Coblentz



Rangahau Mātauranga o Aotearoa | New Zealand Council for Educational Research
Te Pakokori
Level 4, 10 Brandon St
Wellington
New Zealand

www.nzcer.org.nz

<https://doi.org/10.18296/rep.0107>

© New Zealand Council for Educational Research, 2026

PAT trends in Aotearoa New Zealand

**Evidence in Mathematics,
Reading Comprehension,
and Reading Vocabulary
from 2012 to 2026**

David Coblentz

2026

Contents

Acknowledgements	vii
Executive summary	1
Important considerations	1
Summary of findings	2
<i>Mathematics</i>	2
<i>Reading Comprehension</i>	3
<i>Reading Vocabulary</i>	3
<i>Other important findings</i>	3
<i>Implications for recent policy changes</i>	5
<i>Overall picture</i>	5
1. Introduction	6
National achievement landscape	7
International patterns	8
<i>Technology</i>	9
<i>Gender</i>	10
<i>Social and cultural position</i>	10
<i>COVID-19</i>	11
<i>Changing policy landscape</i>	11
<i>Summary</i>	12
Progressive Achievement Tests in Aotearoa New Zealand	13
<i>Scale score</i>	14
2. Methodology	16
Analysis population and model-frame construction	16
Assessment refreshes, linking, and visual alignment	17
School equity measures	18
Ethnicity presentation	19
Model specification and statistical estimation	19
Distributional analysis	20
Year 4 entry and transition analysis	21
Sensitivity analyses and limitations	21
Software used in this report	22
How to read the graphs	22
3. Overall trends	23
4. Year-level patterns	27
5. Year 4 entry and transition analysis	30
5.1 How to read the relative learning figures	30
5.2 Mathematics	31
5.3 Reading Comprehension	32
5.4 Reading Vocabulary	33
5.5 What the transition evidence adds	33

6. Changes across the score distribution	35
7. Gender patterns	37
8. School socioeconomic context (simulated EQI)	42
9. Ethnicity patterns	44
9.1 Overview heatmaps	46
9.2 Absolute achievement differences between ethnicity groups	49
9.3 Māori	53
9.4 Pacific	58
9.5 European/Pākehā	64
9.6 North Asian	70
9.7 South, Southeast, and Other Asian	75
9.8 African	80
9.9 Middle Eastern	85
9.10 Latin American	91
9.11 Ethnicity synthesis	95
10. Regional patterns	97
11. Sensitivity and robustness	101
12. Discussion	105
12.1 Three different achievement histories	105
12.2 Mathematics improved broadly, with most of the 2026 difference already present at Year 4 entry	106
12.3 Reading Comprehension decline began early and was not consistently repaired later	107
12.4 Reading Vocabulary improved, but lower rank bands and one transition lagged	108
12.5 Improvement and equity are different questions	108
12.6 What the historical timing and evidence can establish	110
12.7 Plausible explanations and hypotheses	111
12.7.1 <i>Post COVID-19 cohort replacement and recovery</i>	111
12.7.2 <i>Possible contribution from later policy exposure</i>	112
12.7.3 <i>School responses to weaker Year 4 entry</i>	113
12.7.4 <i>Gender shifts</i>	113
12.7.5 <i>Reading Vocabulary and Reading Comprehension divergence</i>	114
12.7.6 <i>Generative AI and wider digital change</i>	115
12.7.7 <i>Migration patterns</i>	115
12.8 What should be explored next	115
13. Conclusion	118
References	119

Tables

Table 1.	Main evidence streams relevant to interpreting PAT longitudinal trends	8
Table 2a.	PAT series years explored in this report	14
Table 2b.	Average 12-month progress in PAT scale-score units	14
Table 3.	Primary model-frame coverage by assessment series	17
Table 4.	Year 4 entry and transition summary	30
Table 5.	Gender change summary, by assessment series	37
Table 6.	Endpoint boys–girls adjusted gap by year level	41
Table 7.	Retained assessment records, by ethnic group, assessment measure, and year	45
Table 8.	Ethnic group overview: terminal-year within-series heatmap values	49
Table 9.	Māori companion summary	53
Table 10.	Pacific companion summary	59
Table 11.	European/Pākehā companion summary	64
Table 12.	North Asian companion summary	70
Table 13.	South, Southeast, and Other Asian companion summary	75
Table 14.	African companion summary	81
Table 15.	Middle Eastern companion summary	86
Table 16.	Latin American companion summary	91
Table 17.	Terminal-year regional within-series heatmap values	97
Table 18.	Total retained assessment record counts by region and assessment series	98

Figures

Figure 1a.	PAT Mathematics overall national trend with refreshed segment visually aligned	23
Figure 1b.	PAT Reading Comprehension overall national trend with refreshed segment visually aligned	24
Figure 1c.	PAT Reading Vocabulary overall national trend	25
Figure 2a.	Adjusted PAT Mathematics score change from the relevant within-series reference, by assessment year and year level	27
Figure 2b.	Adjusted PAT Reading Comprehension score change from the relevant within-series reference, by assessment year and year level	28
Figure 2c.	Adjusted PAT Reading Vocabulary score change from the relevant within-series reference, by assessment year and year level	28
Figure 3a.	PAT Mathematics Year 4 entry score and later annual interval comparisons for all students	31
Figure 3b.	PAT Reading Comprehension Year 4 entry score and later annual interval comparisons for all students	32
Figure 3c.	PAT Reading Vocabulary Year 4 entry score and later annual interval comparisons for all students	33
Figure 4.	Adjusted score change, by rank band, Years 4 to 8 combined	35
Figure 5a.	PAT Mathematics adjusted gender trajectories relative to the all-students baseline	38
Figure 5b.	PAT Reading Comprehension adjusted gender trajectories relative to the all-students baseline	39
Figure 5c.	PAT Reading Vocabulary adjusted gender trajectories relative to the all-students baseline	40
Figure 6a.	Ethnic group heatmap for PAT Mathematics	46
Figure 6b.	Ethnic group heatmap for PAT Reading Comprehension	47
Figure 6c.	Ethnic group heatmap for PAT Reading Vocabulary	48
Figure 6d.	Modelled ethnic group absolute scale-score trajectories for PAT Mathematics	50
Figure 6e.	Modelled ethnic group absolute scale-score trajectories for PAT Reading Comprehension	51
Figure 6f.	Modelled ethnic group absolute scale-score trajectories for PAT Reading Vocabulary	52
Figure 7a.	Māori Mathematics trajectory	54
Figure 7b.	Māori relative to not-Māori same-starting-mix difference-over-time for PAT Mathematics	55
Figure 7c.	Māori Reading Comprehension focal trajectory	56
Figure 7d.	Māori relative to non-Māori common-reference difference-over-time for PAT Reading Comprehension	56
Figure 7e.	Māori Reading Vocabulary focal trajectory	57
Figure 7f.	Māori relative to non-Māori common-reference difference-over-time for PAT Reading Vocabulary	58
Figure 8a.	Pacific Mathematics focal trajectory	60
Figure 8b.	Pacific relative to non-Pacific common-reference difference-over-time for PAT Mathematics	60
Figure 8c.	Pacific Reading Comprehension focal trajectory	61
Figure 8d.	Pacific relative to not-Pacific same-starting-mix difference-over-time for PAT Reading Comprehension	62

Figure 8e.	Pacific Reading Vocabulary focal trajectory	63
Figure 8f.	Pacific relative to non-Pacific common-reference difference-over-time for PAT Reading Vocabulary	63
Figure 9a.	European/Pākehā Mathematics focal trajectory	65
Figure 9b.	European/Pākehā relative to not European/Pākehā common-reference difference-over-time for PAT Mathematics	66
Figure 9c.	European/Pākehā Reading Comprehension focal trajectory	67
Figure 9d.	European/Pākehā relative to not European/Pākehā same-starting-mix difference-over-time for PAT Reading Comprehension	68
Figure 9e.	European/Pākehā Reading Vocabulary focal trajectory	69
Figure 9f.	European/Pākehā relative to not European/Pākehā same-starting-mix difference-over-time for PAT Reading Vocabulary	69
Figure 10a.	North Asian Mathematics focal trajectory	71
Figure 10b.	North Asian relative to non-North Asian same common-reference difference-over-time for PAT Mathematics	71
Figure 10c.	North Asian Reading Comprehension focal trajectory	72
Figure 10d.	North Asian relative to non-North Asian common-reference over time for PAT Reading Comprehension	73
Figure 10e.	North Asian Reading Vocabulary focal trajectory	73
Figure 10f.	North Asian relative to non-North Asian common-reference difference-over-time for PAT Reading Vocabulary	74
Figure 11a.	South, Southeast, and Other Asian Mathematics focal trajectory	75
Figure 11b.	South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Mathematics	76
Figure 11c.	South, Southeast, and Other Asian Reading Comprehension focal trajectory	77
Figure 11d.	South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Reading Comprehension	78
Figure 11e.	South, Southeast, and Other Asian Reading Vocabulary focal trajectory	79
Figure 11f.	South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Reading Vocabulary	80
Figure 12a.	African Mathematics focal trajectory	81
Figure 12b.	African relative to not African common-reference difference-over-time for PAT Mathematics	82
Figure 12c.	African Reading Comprehension focal trajectory	82
Figure 12d.	African relative to non-African common-reference difference-over-time for PAT Reading Comprehension	83
Figure 12e.	African Reading Vocabulary focal trajectory	84
Figure 12f.	African relative to non-African common-reference difference-over-time for PAT Reading Vocabulary	85
Figure 13a.	Middle Eastern Mathematics one-vs-rest focal trajectory	86
Figure 13b.	Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Mathematics	87
Figure 13c.	Middle Eastern Reading Comprehension one-vs-rest focal trajectory	88
Figure 13d.	Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Reading Comprehension	89
Figure 13e.	Middle Eastern Reading Vocabulary focal trajectory	90
Figure 13f.	Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Reading Vocabulary	90
Figure 14a.	Latin American Mathematics one-vs-rest focal trajectory	92
Figure 14b.	Latin American Mathematics focal-vs-not-group gap change	92
Figure 14c.	Latin American Reading Comprehension one-vs-rest focal trajectory	93
Figure 14d.	Latin American Reading Comprehension focal-vs-not-group gap change	94
Figure 14e.	Latin American Reading Vocabulary one-vs-rest focal trajectory	94
Figure 14f.	Latin American Reading Vocabulary focal-vs-not-group gap change	95
Figure 15a.	PAT Mathematics regional heatmap	99
Figure 15b.	PAT Reading Comprehension regional heatmap	99
Figure 15c.	PAT Reading Vocabulary regional heatmap (2012 to 2014 reference)	100
Figure 16a.	PAT Mathematics sensitivity comparison	102
Figure 16b.	PAT Reading Comprehension sensitivity comparison	103
Figure 16c.	PAT Reading Vocabulary sensitivity comparison	104

Acknowledgements

I am deeply grateful to Peter Keegan, Heleen Visser, Charles Darr, Elliot Lawes, and Sue McDowall for the insight and generosity they brought to reviewing this work. I also warmly acknowledge the 647,213 students and 1,306 schools whose data contributed to this report. Their participation in PAT assessment over the past 15 years made these analyses possible.

Executive summary

This report draws on 15 years of Progressive Achievement Test (PAT) data to show how achievement in Mathematics, Reading Comprehension, and Reading Vocabulary changed from 2012 to 2026. It covers Term 1 assessments for students in Years 4 to 8 attending English-medium state and state-integrated schools. Together, the analyses include just over 3 million assessment records from 647,213 students in 1,306 schools. These comprise almost 1.3 million Mathematics assessments, almost 1.2 million Reading Comprehension assessments, and more than 550,000 Reading Vocabulary assessments.

PAT makes a valuable contribution to national achievement monitoring because it provides a large body of evidence every year through ordinary school use. National and international studies are based on carefully designed representative samples and provide important estimates of achievement at selected points in time. PAT complements these studies by drawing on much larger numbers of students and schools, providing detailed evidence about how achievement patterns vary across time, groups, and parts of the school system.

Important considerations

Five considerations are important when interpreting the findings in this report.

First, two of the three assessments used in this report were redesigned during the period covered here. As a convenient shorthand, this report refers to the earlier versions as “legacy” and the redesigned versions “refreshed”. Mathematics is divided between the legacy PAT Mathematics series from 2012 to 2023 and the refreshed PAT Pāngarau | PAT Mathematics series from 2024 to 2026. Reading Comprehension is likewise divided between the legacy PAT Reading Comprehension series spanning 2012 to 2024 and the refreshed PAT Pānui | PAT Reading Comprehension series running from 2025 to 2026. Reading Vocabulary was not redesigned and forms one continuous series from 2012 to 2026.

The refreshed assessments use broadly comparable PAT scales to the legacy assessments, but they do not necessarily perfectly align at the transition point. For this reason, the report measures trends only within each assessment series. The Mathematics rise described in this report from 2024 to 2026, for example, compares 3 years of the refreshed assessment but makes no claim about change between 2023 and 2024.

Second, because the assessments were taken in Term 1, each year’s results mainly reflect learning during the previous school year. The 2026 results, for example, mostly reflect learning during 2025. This is important to keep in mind throughout the report.

Third, the PAT data were not collected from a nationally representative sample. The findings therefore describe the large population of students and schools using PAT rather than providing exact national estimates. PAT’s particular strengths are its scale, annual timing, and level of detail. Curriculum Insights provides representative national monitoring, while international studies provide broader comparisons. These sources answer partly different questions and are most informative when considered together.

Fourth, the results are based on statistical modelling that adjusts for changes over time in the mix of students and schools using PAT. Annual results are estimated as though they included the same mix of year levels, genders, ethnic groups, and school socioeconomic contexts. The modelling also accounts for repeated assessments of students and similarities among students attending the same school as well as school socioeconomic context.

Finally, PAT scale-score points are the measurement units used to compare achievement within each subject throughout the report. The Mathematics and two reading assessments have different scales, so their scale points should not be compared directly with one another. To make the size of a change easier to understand, this summary also compares it with the progress students typically make during one school year. These findings describe changes in average achievement across large groups of students. They do not mean that every individual student's achievement changed by the reported amount.

Summary of findings

The clearest overall message of this report is that Mathematics, Reading Comprehension, and Reading Vocabulary changed in different ways over time. The PAT record does not support a single story of general decline or general recovery. Mathematics declined and then improved. Reading Comprehension showed a longer and more sustained decline. Reading Vocabulary followed a more positive course.

Mathematics

PAT Mathematics was broadly stable through most of the 2010s. Scores began to decline in 2019, before COVID-19, and reached their lowest point in the legacy series in 2022, reflecting reduced COVID-19 era learning in 2021. By 2023, most of that decline had been recovered, although average scores remained slightly below their 2012 level. Within the refreshed assessment, Mathematics scores rose in both 2025 and 2026. By 2026, the adjusted average was 1.64 PAT scale points above 2024. This corresponds to about one-third of the progress students typically make during a single year of schooling.

The recent improvement was broad. The positive direction appeared at every assessed year level, in every geographic region included in the analysis, and in every ethnic group. Improvement was also visible through the lower, middle, and upper parts of the score distribution, although lower-scoring Year 8 students shared less fully in the gain. The broad increase was therefore not produced by an isolated region or a small part of the student population.

The most important age pattern is that much of the recent increase in Mathematics scores was already present when students reached Year 4, when most students first take PAT assessments. In 2026, the adjusted Year 4 average was 2.36 PAT scale points above 2024. This is equivalent to almost half of the progress students typically make in a year. The four later transition comparisons were also positive, but smaller, ranging from 0.18 to 0.26 points.

The recent national increase in Mathematics was led by several successively stronger groups of students reaching Year 4, with smaller favourable movements added as students progressed through the later year levels. Those later differences are modest in any single transition, but they could become important if they continue across successive years. Future results will show whether the stronger 2026 Year 4 group retains its advantage as it moves through school and whether later groups also enter Year 4 at a similarly strong level.

Reading Comprehension

Reading Comprehension scores began declining before COVID-19 and continued to fall through 2024, the final legacy year. Scores were already declining before COVID-19 and continued to fall through 2024, the final year of the earlier assessment. By then, the adjusted average was 1.76 PAT scale points below 2012, equivalent to about one-quarter of the progress students typically make in a school year between Years 4 and 8. The decline appeared at every assessed year level, in every region included in the analysis, throughout the score distribution, and in almost every ethnic group. Alternative ways of defining the students and schools included in the analysis changed the estimated size of the decline, but all showed the same downward direction.

Much of this decline was already visible by the time PAT assessments begin in Year 4. From 2015 onward, successive Year 4 groups generally entered the PAT testing window with lower average adjusted scores than their predecessors had in 2012. Between 2020 and 2024, Year 4 scores were about two points lower, equivalent to roughly one-quarter of the progress usually made in a year. The modelled movement from Year 4 to Year 5 became more favourable over time and partly offset that weaker starting position. This partial catch-up, however, was not consistently maintained through Years 5 to 7. At Year 8, the upper part of the score distribution declined more over time than the middle. The longer Reading Comprehension decline therefore began early for many groups of students and was not reliably repaired later.

The refreshed Reading Comprehension assessment showed a rise from 2025 to 2026. The increase of 0.51 scale points was less than one-tenth of typical annual progress. The direction was positive across ethnic groups and regions, but not equally clear at every year level. Year 7 improved most, while Year 8 changed very little. This is an encouraging result, but it does not yet establish that the longer decline has ended. The next results will show whether the improvement continues and whether it reaches older students.

Reading Vocabulary

Reading Vocabulary followed a positive path. The adjusted average in 2026 was 1.34 scale points above 2012, equivalent to about one-fifth of typical annual progress. Every assessed year level, every region, and every ethnic group ended above its own position in 2012. The improvement was nevertheless uneven. Movement from Year 4 to Year 5 was consistently strong, while movement from Year 6 to Year 7 remained weak. Students in the lower part of the score distribution did not decline over time, but they gained less than a quarter as much as those in the rest of the distribution.

Other important findings

The growing separation between Reading Vocabulary and Reading Comprehension is an important finding in this report. Vocabulary scores rose while Comprehension scores declined through the earlier assessment period. The same pattern appeared among students who took both assessments in the same year and at the same year level. For these students, the adjusted Vocabulary score relative to the Comprehension score widened by 2.28 points between 2012 and 2024. The divergence therefore cannot be explained simply by different students taking the two assessments.

Across all three assessments, the difference between boys' and girls' scores moved towards boys, but it did so in different ways. In Mathematics, girls lost more ground during the earlier period and gained less since the pandemic-era disruptions. Overall, the difference moved 2.42 scale points towards

boys between 2012 and 2023, equivalent to about half a year of typical Mathematics progress. It then moved another 0.87 points in the same direction from 2024 to 2026 in the refreshed series.

In Reading Comprehension, girls scored higher than boys throughout the 15 years of this study, but both groups declined overall and girls declined more than boys. For Reading Vocabulary, average score changes over time came mainly from rising scores among boys, while girls' scores changed little. By 2026, average Vocabulary scores for boys and girls were very similar, with boys slightly higher overall.

There are two related ways in which equity among ethnic groups is explored in this report. One is the size of achievement differences between groups at any given point in time. The other is how each group has changed over time, both from its own earlier position and compared with how much other groups changed. These measures can tell different stories. A group can improve while a large achievement gap remains, and a gap can narrow because a higher-scoring group declines rather than because a lower-scoring group improves.

Large, adjusted differences remained between ethnic groups throughout the period covered in this report. In Mathematics in 2026, the full spread was 11.7 scale points, roughly 2 years of typical progress. North Asian (Chinese, Japanese, and Korean) students were about 7.1 points above the next closest group. Corresponding differences were 4.2 points in Reading Comprehension and 4.4 points in Reading Vocabulary, equivalent to a little over half a year of typical progress in each of these subjects. Recent improvement among many groups therefore has coexisted with larger and more stable differences in achievement between groups.

School socioeconomic context was associated with even larger differences. Across the three subjects, students in schools serving communities with greater socioeconomic barriers scored about 10 scale points below students in schools serving communities with fewer barriers after accounting for the other characteristics in the models. In every subject, this was more than 1 year of typical academic progress as measured by PAT assessments.

Some of these differences narrowed over time, but narrowing was not always good news. In the earlier Mathematics assessment, scores fell in the lower barrier context while remaining broadly stable in the higher barrier context. In Reading Comprehension, both contexts declined, but the lower barrier context declined more. Reading Vocabulary followed the more positive pattern, with improvement in the higher barrier context. A smaller gap is not an improvement in equity when it is produced by falling achievement among students who previously scored higher.

The same broad patterns remained across four alternative analyses. Three used narrower samples, focusing on students tested in more than 1 year, schools that used PAT consistently, or schools that used PAT every year from 2020 to 2026. The fourth allowed each school to follow its own trend over time. In all of them, Mathematics still reached its low point in 2022 and rebounded in 2023. Reading Comprehension still ended the earlier assessment period below its starting level, although the estimated decline was smaller when schools were allowed to follow different trends. Reading Vocabulary still ended positively. These checks are intended to strengthen confidence that the main direction and timing are not simply products of changing PAT participation or one set of modelling choices.

Implications for recent policy changes

The Mathematics rebound began in 2023, before the daily hour requirement introduced in 2024 and before the revised Mathematics curriculum became compulsory in 2025. Those changes cannot explain the beginning of the recovery, although they may have helped sustain it. The 2026 Mathematics results are the first to follow a full school year under the revised curriculum, so later results will be more informative about whether improvement continues. For Reading Comprehension, the refreshed assessment began in 2025, leaving less opportunity to separate policy effects from the assessment change. It is important to keep in mind that, while PAT can show when achievement changed, this analysis cannot identify the effect of any single policy or teaching approach.

Overall picture

Aotearoa New Zealand benefits from having several credible sources of achievement evidence. PAT provides annual detail from ordinary school use. Curriculum Insights provides representative national monitoring linked to the New Zealand Curriculum. PISA provides a long-term view of applied reading and mathematics among 15-year-olds, while TIMSS and PIRLS provide international comparisons closer to the primary years. These studies can show whether PAT patterns also appear in independently designed assessments and whether changes in Aotearoa New Zealand resemble those occurring elsewhere. When different sources point in the same direction, confidence in a wider system change increases. When they differ, the difference can reveal that subjects, ages, assessments, or time periods are moving differently.

Taken together, the PAT record shows how achievement differences build up across schooling. Stronger or weaker positions are often already visible by Year 4. The recent Mathematics improvement is broad and encouraging, but much depends on whether the stronger 2026 Year 4 group retains its advantage. Reading Comprehension shows a longer decline beginning early, with only a first small sign of improvement. Reading Vocabulary has risen, but students in the lower part of the score distribution have shared much less in that gain. Future monitoring should follow these groups through later years, pay close attention to the Year 6 to Year 7 transition, and test whether improvement becomes more sustained and more evenly shared.

1. Introduction

Long-term trends in student achievement are bound up with some of the most persistent questions in education. They help shape claims about whether the system is improving or declining. They influence curriculum debates, assessment policy, resource allocation, school improvement priorities, public confidence in schooling, and many other aspects of education. They also inform broader debates about equity, language, culture, opportunity, and the extent to which schools mitigate or reproduce wider social inequalities. If insufficient longitudinal data are available about student achievement, the public conversation risks becoming overly reactive to sporadic or isolated data.

Overall, Aotearoa New Zealand has a comparatively rich monitoring environment in English-medium schools. International studies such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literacy Study (PIRLS) have provided repeated evidence about Aotearoa New Zealand student achievement within globally standardised assessment programmes. National monitoring studies, including the National Monitoring Study of Student Achievement (NMSSA) and its more recent successor, the Curriculum Insights and Progress Study, provide carefully sampled curriculum-linked evidence at selected year levels. School-use assessment systems, including the New Zealand Council for Educational Research's (NZCER's) PATs, provide another kind of evidence, repeated assessment information generated through the ordinary work of large numbers of students and schools across Aotearoa New Zealand. These evidence streams overlap in the broad questions they inform, but they differ in many details, such as sampling frame, method of administration, and frequency of collection.

That diversity is both a strength and a challenge for anyone who wants to understand the long-term direction of student achievement in Aotearoa New Zealand. It is a strength because multiple monitoring streams can be brought to bear on the same broad questions. It is a challenge because public debate often wants a simple answer to the question of whether achievement is rising or falling. In practice, the necessary question is rather how to interpret a layered evidence system in which different tools observe different parts of schooling at different intervals for different purposes. Convergence between evidence streams matters, but divergence also matters. Divergence may reflect differences in measurement and sampling, but it may also be a consequence of different timing or trajectories across subsets of learners.

PAT Mathematics, PAT Reading Comprehension, and PAT Reading Vocabulary provide annual evidence from 2012 to 2026 for the wider discussion about student achievement in Aotearoa New Zealand. In addition to the periodic evidence provided by PISA, TIMSS, PIRLS, NMSSA, and Curriculum Insights, PAT assessments have the potential to show annual trends within the same assessment family, based on a large volume of achievement data that spans much of the country. That matters in a policy environment where long-term trends are often inferred from periodic snapshot studies. While cyclic studies can show where the system stands at important points in time, they are less well suited to showing how achievement moved between those points or whether change was gradual or abrupt. PAT can help fill that gap, provided its context and interpretive limits are kept in mind.

National achievement landscape

In 2022, the Ministry of Education (the Ministry) drew together available English-medium evidence on literacy and mathematics in Aotearoa New Zealand. The literacy overview describes a mixed picture. Many students perform well and make expected progress, but achievement and progress vary widely, and some indicators have declined over time (Caygill et al., 2022). The mathematics overview also presents a mixed picture. While many students are meeting expectations, progress varies substantially and access to key learning opportunities is uneven (Caygill & Zhao, 2022). These overview papers help explain why long-term trends remain contested. The existing evidence does not cleanly support a simple account in which everything is improving or everything is declining.

PISA is often the most visible international study in Aotearoa New Zealand discussions of student achievement. It has assessed and reported on 15-year-olds' reading, mathematics, and science literacy since 2000. PISA assesses students' ability to apply knowledge and skills to real-life challenges. The New Zealand PISA 2018 summary reported that Aotearoa New Zealand remained above the OECD average in reading, mathematics, and science, but that the long-term trend had declined in all three domains (May et al., 2019). The Aotearoa New Zealand PISA 2022 summary reported a 15-point fall in mathematics from 2018 to 2022, matching the OECD average fall over the same period. Reading and science were broadly similar to 2018, and the proportion of low achievers in mathematics increased (May & Medina, 2023). The Ministry also published a non-response bias analysis because Aotearoa New Zealand did not meet the strict international participation standards for the 2022 cycle (May et al., 2023).

TIMSS and PIRLS provide other international points of reference. TIMSS reports mathematics and science at Year 5 and Year 9. The 2023 Year 5 mathematics report found that Aotearoa New Zealand average achievement was higher than in 1994 but similar to recent cycles. It also reported a wide spread of achievement, higher Year 5 mathematics scores for boys, and large differences by school socioeconomic context (Caygill et al., 2024). The 2023 Year 9 mathematics report found average achievement lower than in 1994 but broadly similar to recent cycles (Caygill & Borthwick, 2024). PIRLS reports reading literacy at Year 5. Aotearoa New Zealand's PIRLS 2021 report found no significant change from 2016 to 2021 in the mean score, score distribution, or proportions reaching PIRLS reading benchmarks, although Aotearoa New Zealand's relative standing improved, partly because many other countries declined (Chamberlain & Forkert, 2023). This contrasts with the decline in PAT Reading Comprehension over much of the same period, although the studies assess different constructs, year levels, and student populations.

NMSSA assessed Year 4 and Year 8 achievement across learning areas in English-medium state and state-integrated schools. The 2022 NMSSA mathematics and statistics report provided nationally representative evidence for about 2,000 students at each of Year 4 and Year 8 (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2023). The successor to NMSSA, Curriculum Insights, has begun to reshape the national monitoring landscape by moving to an annual system of assessing reading, writing, and maths in addition to providing cyclic snapshots of each learning area every few years. The 2023 mathematics and reading report was the first year of the study, and, from 2025, reading, writing, and mathematics have been assessed and reported on every year (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2024a). The 2023 Curriculum Insights reading report estimated that about half of students were meeting the expected curriculum reading level for their age. It also found that a substantial minority were more than a year behind the benchmark, particularly by Year 8 (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2024b).

PISA, TIMSS, PIRLS, NMSSA, and Curriculum Insights are all important monitoring tools. The primary addition that PAT can provide to the existing conversation about student achievement in Aotearoa New Zealand is high-volume, annual same-family evidence across the later primary school years.

TABLE 1. Main evidence streams relevant to interpreting PAT longitudinal trends

Evidence stream	Main population or year levels	Cycle	Main contribution to this report	Limitations
PAT	School-use assessment data across Years 4 to 8	Annual school-use evidence	Main empirical basis for the three longitudinal legacy series	School-use data, not a national probability sample series
PISA	15-year-olds. Reading, mathematics, and science literacy	Three-year international cycle, with disruption to the 2022 timing	Long-run and international context for achievement and equity debate	Not Years 4 to 8. Age-based. Scale not numerically comparable with PAT
TIMSS	Year 5 and Year 9 mathematics and science	Four-year international cycle	International mathematics context closer to the primary and lower-secondary years	Framework, sampling design, and scale differ from PAT
PIRLS	Year 5 reading literacy	Five-year international cycle	International reading context near middle primary schooling	Reading construct, cycle, and scale differ from PAT
NMSSA	Year 4 and Year 8 curriculum monitoring	Five-year cycle	Nationally representative curriculum-linked snapshots at key year levels	Snapshot evidence at selected levels rather than annual same-family middle-years series
Curriculum Insights	Annual monitoring of reading, writing, and mathematics at Years 3, 6, and 8	New national monitoring architecture. Annual foundational-area assessment from 2025	Current and future national curriculum monitoring context	Recent, so cannot provide annual data historically

International patterns

Internationally, by the late 2010s, several large-scale achievement studies already pointed to stalled improvement or outright declines in parts of the OECD and in some other high-income systems. The pandemic then arrived on top of those existing patterns.

PISA is one of the clearest sources for a longer view of global student achievement, with the 2018 OECD report noting that most participating education systems had not shown sustained improvement across reading, mathematics, or science over the years of their PISA participation. For countries with long-term data, the average OECD reading trend had improved slowly up to around 2012 and then begun declining. PISA 2018 was consequently already being read internationally as evidence of stagnation or decline in many education systems before the COVID-19 disruption (OECD, 2019a). Aotearoa New Zealand's PISA 2018 report followed this pattern, as it reported long-term declines in

reading, mathematics, and science, despite Aotearoa New Zealand remaining above the OECD average (May et al., 2019).

However, other international evidence gives a more complex picture. TIMSS and PIRLS both show that change has not been the same across subjects, ages, or countries. PIRLS 2021 reported reading achievement during a cycle affected by pandemic conditions, but it also continued a longer reading-literacy trend series. TIMSS 2023 showed that mathematics performance and gender patterns varied across countries and year levels, with some systems showing stability and others showing larger changes (von Davier et al., 2024). These studies are useful because they caution against reading the international context as a single story of decline. They also foreshadow the finding discussed in detail later that PAT Reading Comprehension, PAT Reading Vocabulary, and PAT Mathematics trends in Aotearoa New Zealand are not the same.

Technology

During the period covered by this report, children became more heavily engaged with digital environments both in and outside of school. That shift matters for Reading Comprehension and Reading Vocabulary, but the evidence does not support a simple claim that digital technology either improves or damages learning on its own. OECD analysis based on PISA 2012 found that heavy investment in school Information and Communications Technology (ICT) had not produced clear improvements in reading, mathematics, or science performance, and that very frequent computer use at school was associated with weaker outcomes after background adjustment (OECD, 2015). The 2023 UNESCO Global Education Monitoring Report reached a similarly cautious conclusion. It described technology as potentially useful but argued that strong independent evidence of educational benefit remains limited and that access, governance, and teacher support help shape whether technology ends up benefiting learners (UNESCO, 2023).

PISA 2018 showed that reading for enjoyment had declined across the OECD from 2009 to 2018, but that girls continued to report more enjoyment of reading than boys (OECD, 2019b). An Aotearoa New Zealand PISA 2018 reading report described a similar pattern here, with less reading for enjoyment and increases in online reading activities (Medina & McGregor, 2019). Research comparing print and digital reading also points to a shift in reading conditions. A meta-analysis by Delgado and colleagues found that, across studies from 2000 to 2017, paper-based reading was associated with stronger comprehension than screen-based reading, especially for informational texts and when under time pressure (Delgado et al., 2018). A later meta-analysis of leisure digital reading found that digital reading habits did not appear to support text comprehension in the same way as traditional print reading (Altamura et al., 2025).

In contrast, technology has a more nuanced relationship with vocabulary acquisition. Vocabulary grows through repeated exposure to language across a wide range of contexts (Mol & Bus, 2011; Rowe, 2012; Torppa et al., 2022). Digital media can add to these language encounters, but its effects depend less on time spent consuming media than on content quality and context. A recent meta-analysis found a small positive association between screen media and vocabulary for young children, with stronger effects for educational content, e-books, and interactive media than for general screen exposure (Jing et al., 2023). Well-designed video and digital books can support vocabulary learning (Egert et al., 2022; Gowenlock et al., 2024). Joint use by adults and children has a small positive association with young children's learning from digital media, although the evidence base remains limited (Taylor et al., 2024). PAT Reading Vocabulary trends should therefore be interpreted as likely reflecting a wider mix of language experiences, not necessarily just reading exposure alone.

Gender

Gender is a useful but limited lens for interpreting achievement trends. Score ranges for boys and girls overlap substantially, and research on gender differences in education often cautions against deterministic interpretations (Hyde et al., 2019; Reilly, 2020).

International evidence related to gender differences is most consistent for reading comprehension. Girls have generally outperformed boys in reading in large-scale international assessments. In PISA 2022, girls scored higher than boys in reading in nearly all participating systems; in Aotearoa New Zealand, girls outperformed boys in reading by 26 score points, while boys scored higher than girls in mathematics by 10 score points (OECD, 2023). PIRLS 2021 similarly reported higher average reading achievement for girls in most participating countries. Aotearoa New Zealand was not an exception, as the Year 5 gender difference was 19 score points in favour of girls, though this was the lowest Aotearoa New Zealand PIRLS gender gap in 20 years (Chamberlain & Forkert, 2023). Academic evidence is generally consistent with these broad patterns. One study, for example, found gender differences in reading and larger differences in writing across several decades of National Assessment of Educational Progress (NAEP) data (Reilly et al., 2019). Reading motivation also matters, because girls often report more frequent reading and more positive reading attitudes than boys, and reading frequency and print exposure are associated with vocabulary growth, reading comprehension, and broader literacy development (Logan & Johnston, 2009; Mol & Bus, 2011).

Some early-language research reports a small female advantage in vocabulary. For example, Eriksson et al. (2012), drawing on data from more than 10,000 children across 10 language communities, found girls slightly ahead in communicative gestures, productive vocabulary, and early word combination. Experimental work with children aged 5 to 7 also suggests that girls may make stronger use of phonological and semantic familiarity when learning new words (Kaushanskaya et al., 2013). However, not all studies find a universal female advantage in all aspects of language development. Law et al. (2013), using a large UK cohort, found a slight male advantage in receptive vocabulary at school entry, while girls were stronger in early reading.

Analyses of mathematics performance have generally found small average gender differences and large overlap, but also variability across countries (Else-Quest et al., 2010; Lindberg et al., 2010). Recent international monitoring nevertheless gives reason to treat mathematics gender trends as a substantive part of the achievement landscape. TIMSS 2023 reported a Grade 4 mathematics advantage for boys in many participating systems, and Aotearoa New Zealand's Year 5 TIMSS mathematics results showed a statistically significant difference favouring boys for the first time across eight TIMSS cycles (Caygill et al., 2024; von Davier et al., 2024). These patterns highlight the importance of examining whether PAT mathematics trends differ by gender, by year level, and by parts of the achievement distribution.

Social and cultural position

International evidence consistently shows that achievement is patterned according to ethnicity and socioeconomic status, and that learners from minority, Indigenous, low-income, migrant, or non-dominant language communities often face structural barriers that affect their achievement on school assessments. PISA 2018 reported that immigrant-background students had become a larger share of students across OECD countries and that achievement gaps were strongly connected to socioeconomic and language barriers (OECD, 2019b). The 2020 United Nations Educational, Scientific and Cultural Organization (UNESCO) Global Education Monitoring Report described exclusion from

education as linked to identity, language, migration, disability, poverty, and other social positions (UNESCO, 2020).

Aotearoa New Zealand's international results also show that socioeconomic background is an important context for student achievement in Aotearoa New Zealand. PISA 2022 reported a 102 score point mathematics gap between socioeconomically advantaged and disadvantaged students in Aotearoa New Zealand, compared with an OECD average gap of 93 points, and found that socioeconomic status accounted for 16% of variation in mathematics performance, which is close to the OECD average of 15% (OECD, 2023). Curriculum Insights points in the same direction. In its 2023 statistical analyses of reading and mathematics at Years 3, 6, and 8, Schooling Equity Index (EQI) was the strongest school-level predictor in both domains, with greater socioeconomic barriers associated with lower achievement (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2026a).

COVID-19

The pandemic intensified many of the pre-existing patterns described above. International evidence generally shows learning losses during the pandemic period, with substantial variation across countries, subjects, and groups of students. Betthäuser and colleagues' meta-analysis, which included data from many countries, found evidence of learning losses, especially from the first phase of school closures, and found larger losses in mathematics than in reading in much of the available evidence (Betthäuser et al., 2023). A joint report by UNESCO, UIS, UNICEF, the World Bank, and the OECD argued that the pandemic intensified existing learning problems and that countries varied in how far they had moved from emergency response into recovery (OECD et al., 2022). The 2022 PISA results sit in that same frame. While these results are important, they should not be read as if the pandemic were the only relevant explanation for all post-2018 change.

Changing policy landscape

The final years of this analysis overlap with a period of rapid policy change in English-medium schooling. This context matters when recent data are interpreted.

The changes built on the Literacy & Communication and Maths Strategy launched in March 2022. During 2023, the Common Practice Model was developed to provide more consistent, evidence-informed guidance for teaching literacy and mathematics. It was initially scheduled for use from 2024 but was reviewed in 2024 alongside the revised English and Mathematics content. It should therefore be treated as part of the policy development pathway rather than as a discrete intervention implemented uniformly from Term 1 2024 (Education Gazette, 2023; Ministry of Education, 2022, 2024c).

From the start of Term 1 2024, schools with students in Years 0 to 8 were required to provide an average of at least 1 hour a day for each of reading, writing, and mathematics. This established a national minimum for teaching time while allowing schools to organise that time across their programmes (Education Gazette, 2024). The Government also set a target that, by December 2030, 80% of Year 8 students would be at or above the expected curriculum level for their age in reading, writing, and mathematics (Stanford, 2024a).

Reading policy was moved toward a structured literacy approach. In May 2024, the Government announced that structured literacy would be required in all state schools from Term 1 2025, supported by professional learning, books and resources, phonics checks, and additional help for students who needed it (Stanford, 2024b). Phonics checks were trialled in Term 4 2024 and made available in 2025.

From Term 1 2026, schools were required to administer them after approximately 20 and 40 weeks of a student's first year at school (Ministry of Education, 2024a, 2026c).

Mathematics policy changed on a similar timetable. The Make It Count action plan was announced in August 2024, bringing forward revised Years 0 to 8 mathematics content to Term 1 2025 and providing national resources and professional learning (Luxon & Stanford, 2024). Revised Years 0 to 6 English content and Years 0 to 8 mathematics content were published in October 2024 for required use during 2025 (Ministry of Education, 2024b). Updated Years 0 to 10 English and Mathematics and Statistics content was released in October 2025 for required use from Term 1 2026 (Ministry of Education, 2025c).

Assessment and reporting also became more standardised. In July 2024, the Ministry announced a transition to twice-yearly progression monitoring in reading, writing, and mathematics for Years 3 to 8. From 2026, twice-yearly assessment became a requirement, with schools able to use PAT, SMART, or e-asTTle. PAT therefore sits within a new national assessment and reporting environment as well as providing the historical school-use data analysed in this report (Ministry of Education, 2024a, 2025a).

The policy programme continued to develop in 2026. In May, the Ministry released a consolidated Foundational Learning Action Plan that brought together Make It Count, Make It Write, and a new reading action plan called Read to Succeed. The plan includes further assessment checks, teacher support, classroom resources, and targeted interventions in reading, writing, and mathematics (Stanford, 2026). These May 2026 initiatives post-date the Term 1 PAT observations used for 2026 and cannot explain those results.

More generally, the changing education landscape needs to be considered carefully when interpreting the most recent years presented in this report. The 2024 Mathematics refreshed observations predate classroom implementation of the revised 2025 mathematics curriculum. The 2025 Term 1 observations were collected as implementation began and predominantly reflect learning during 2024 and could potentially have captured any effects of the 1-hour daily practice requirement. The 2026 Term 1 observations are the first that can reflect a completed school year under many of the new arrangements. The first refreshed Reading Comprehension point in 2025 likewise provides a largely pre-implementation baseline for the later policy changes. The refreshed Mathematics and Reading Comprehension results can potentially provide early descriptive signals during this transition, but they cannot identify the effect of any single policy or teaching approach.

Summary

The international context and shifting policy landscape sets the scene for reading the Aotearoa New Zealand PAT results which will be discussed in detail later in this report. First, some achievement concerns pre-date COVID-19. Second, the pandemic period is clearly an important part of the story both globally and in Aotearoa New Zealand, as PAT achievement showed clear pandemic-era disruption in the early 2020s, especially in mathematics (Coblentz, 2026). Third, reading comprehension, vocabulary, and mathematics should be expected to move differently because they are shaped by partly different learning conditions. Fourth, gender- and ethnicity-related patterns are part of the global achievement landscape, even though categories and mechanisms differ by country. Fifth, while Aotearoa New Zealand doesn't exist in a vacuum and it is important to place achievement trends in their broader global context, we should expect deviations from global patterns both because the global picture is one of nuance and complexity and because Aotearoa New Zealand exists in its own unique history and context. Finally, during the most recent years, PAT Mathematics and Reading Comprehension have both been refreshed at roughly the same time as curriculum and policy changes that also have the potential to affect student achievement.

Progressive Achievement Tests in Aotearoa New Zealand

Many Aotearoa New Zealand schools use the Progressive Achievement Tests (PATs), a series of standardised, multiple-choice assessments developed by the New Zealand Council for Educational Research (NZCER) and normed for Aotearoa New Zealand students. PATs can be administered to students in Year 4 through Year 10 (roughly ages 8–15). There are PAT assessments for Mathematics (Pāngarau), Reading Comprehension (Pānui), Reading Vocabulary, Listening Comprehension, Punctuation and Grammar, and more recently the English-medium Tuhituhi | Writing. Each test assesses proficiency in its domain. For example, the PAT: Mathematics measures skills in number knowledge, algebra, geometry, and statistics, while the PAT: Reading Comprehension assesses how well students understand and interpret written texts. Schools typically administer these tests once or twice a year to gauge student achievement in core skill areas, with the heaviest use historically being in Term 1.

Importantly, PATs are intended as low-stakes, formative assessments that allow teachers to identify students' strengths and weaknesses, determine what kind of teaching support or materials might be needed, and track progress over time. Because the tests are standardised and nationally norm-referenced, a student's performance can be compared against a national sample to see how they are achieving relative to peers. More importantly, for this study, PATs allow measurement of student progress year to year. As each PAT subject test is scored on a continuous scale, teachers (and researchers) can observe how a student's scale score grows as they advance from one year level to the next. In this way, PAT data can help quantify whether students are making expected gains in learning. Large-scale patterns in PAT scores over multiple years and across many students can also provide an indicator of overall learning trends in Aotearoa New Zealand.

PAT assessments have a long history in Aotearoa New Zealand education. NZCER first developed the tests in the 1960s and has revised them over time. They remain one of the main sets of tests used by Aotearoa New Zealand schools (NZCER Assessment Services, 2024).

In recent years, NZCER has continued to refresh PAT assessments so they better reflect the needs of learners and schools in Aotearoa New Zealand. The refresh aimed to make the assessments more equitable and accessible. It also aimed to strengthen their cultural relevance and authenticity. The refreshed PAT Mathematics series, renamed PAT Pāngarau | PAT Mathematics, replaced previous versions from 2024, as did revised PAT Pānui | PAT Reading Comprehension assessments from 2025. PAT Reading Vocabulary was not part of the 2023 to 2024 Pānui refresh and remains a separate assessment (New Zealand Council for Educational Research, 2024b). The refresh may also have changed student engagement with the assessments, but that possibility is not tested here. These refreshes create measurement boundaries within the Mathematics and Reading Comprehension time series. The analysis does not attempt to formally bridge the legacy and refreshed series within each subject and instead models each series separately. Selected plots visually align the first refreshed point with the final legacy point so that movement within each series can be seen simultaneously, but this is a display convention and change across the refresh boundary is not formally estimated.

However, although this analysis did not apply a formal bridging model, the legacy and refreshed assessments were linked through Rasch modelling during their development and trialling, with the aim of maintaining a broadly comparable PAT scale. It is therefore reasonable to regard scale-score units in the two series as broadly comparable. The linking process does not guarantee exact numerical continuity, however, so an observed difference between the final legacy year and the first refreshed year should not automatically be interpreted as a change in student achievement. Accordingly, this report bases its formal trend conclusions on changes within each series and does not estimate change across the refresh boundary.

Table 2a. PAT series years explored in this report

PAT measure	Legacy series	Refreshed series	Treatment in this report
PAT Mathematics	2012 to 2023	2024 to 2026	The series are modelled separately. Refreshed results are visually aligned in selected plots. Cross-boundary change is not estimated.
PAT Reading Comprehension	2012 to 2024	2025 to 2026	The series are modelled separately. Refreshed results are visually aligned in selected plots. Cross-boundary change is not estimated.
PAT Reading Vocabulary	2012 to 2026	Not applicable	Treated as a single continuous assessment series.

Scale score

In PAT, a scale score locates a student’s estimated achievement on a measurement scale for a particular learning area. Each subject has its own scale, but different test forms and year levels within that subject are calibrated to the same scale, allowing achievement and progress to be compared over time. A student’s responses are used to estimate their scale location through a Rasch measurement model that takes account of item difficulty (Darr, 2024; New Zealand Council for Educational Research, 2024a). The resulting scale score indicates the student’s estimated achievement relative to the calibrated difficulty of the PAT questions. Because the scale is designed to have equal intervals, a 5-point difference represents approximately the same distance on that subject’s scale wherever it occurs. A common scale supports comparison over time, but it is distinct from curriculum levels. Individual scale scores should be interpreted with their reported margin of error (NZCER Assessment Services, 2024).

PAT Mathematics, PAT Reading Comprehension, and PAT Reading Vocabulary each have their own scale, so scale-score changes should be compared within, rather than between, subjects.

Across Years 4 to 8, average annual progress is usually several scale-score units, but it varies by subject and year. This means that many of the annual changes reported in this study are smaller than a typical year-level difference, even when they are statistically clear in a large dataset (New Zealand Council for Educational Research, 2025).

TABLE 2b. Average 12-month progress in PAT scale-score units

Transition	Mathematics	Reading Comprehension	Reading Vocabulary
Year 4 to Year 5	7.2	9.0	8.5
Year 5 to Year 6	6.1	8.0	7.8
Year 6 to Year 7	3.3	7.6	6.3
Year 7 to Year 8	4.5	6.6	5.1

A raw-score example is also useful. PATs are commonly experienced by students as multiple-choice tests, and NZCER explains that raw scores such as 22 out of 40 are converted onto the relevant scale while accounting for item difficulty (NZCER Assessment Services, 2024). Around the middle of a typical Year 6 test, one scale-score point is roughly comparable with about one raw-score question, approximately 1.2 questions in the Mathematics example used here. A two-point scale-score difference would therefore often be in the order of two to three questions on a 40-item test.

The scale-score changes reported here should therefore generally not be interpreted as large changes in the capability of a typical individual student. A difference of around two scale-score points is modest at the individual level, and frequently the amount of change discussed in this report is smaller still, often only a fraction of a scale-score point. However, when shifts of this size are observed as an average across many thousands of students, they can nonetheless become meaningful as a system-level indicator, particularly when they are part of a long-term trend.

The following section defines the analysis population and explains how the assessment series, scale scores, and contextual variables were prepared and modelled.

2. Methodology

This section defines the analysis population and assessment series, describes how the PAT datasets were prepared, explains how school socioeconomic context was represented across the transition from school decile to EQI, and sets out the statistical modelling approach. The report analyses PAT Mathematics, PAT Reading Comprehension, and PAT Reading Vocabulary results from 2012 to 2026, using the legacy and refreshed assessment segments shown in Table 2a. The analyses examine national, year-level, distributional, gender, ethnicity, regional, and Year 4 entry and transition patterns, together with a series of sensitivity analyses.

Analysis population and model-frame construction

The model frames retained eligible Term 1 assessments for students in Years 4 to 8 attending English-medium state and state-integrated schools. Years 9–10, private schools, kaupapa Māori and Māori-medium kura, and other non-English-medium settings were outside the study population. Restricting the analysis to Term 1 and Years 4 to 8 improves comparability because PAT use has historically been concentrated in Term 1 and secondary school use patterns sometimes differ, particularly in Year 10.

While PAT assessments have been used in Aotearoa New Zealand for decades, electronic storage of these assessments began in 2010. Because 2010 and 2011 had lower online data volume than subsequent years, 2012 is used as the starting date for the longitudinal analyses.

Table 3 shows the primary model frames. These are large assessment datasets. A cohort-based calculation using the Ministry's annual roll series and 2025 national projection data gives an approximate total of 1.2 million learners who passed through at least one of Years 4 to 8 between 2012 and 2026 (Ministry of Education, 2025b, 2026b). Across this period, the primary model frames include at least one eligible Mathematics assessment for approximately 50% of these learners, a Reading Comprehension assessment for approximately 45%, and a Reading Vocabulary assessment for approximately 25%.

All analyses were restricted to students attending English-medium, state, and state-integrated schools to align the assessments used with the curriculum they are designed to represent as much as possible. PATs are developed in the context of the *The New Zealand Curriculum (NZC)* (Ministry of Education 2007), and they have historically been used primarily in these schools. The tests are available in English only and are used infrequently in non-English-medium settings. In addition, the EQI is defined and published only for state and state-integrated schools and kura (Ministry of Education, 2020). Private schools fall outside the EQI framework, and many do not follow NZC in full, making them less comparable with state and state-integrated schools.

The PAT Pāngarau | PAT Mathematics assessment has adaptive and static test versions. Adaptive tests have historically been used less than static tests and were excluded from all analyses. In addition, alternative test versions are available for schools for a variety of purposes, including testing Year 3 students toward the end of the school year and providing additional assessments for retesting students. These assessments provide only a small amount of the total data available and are not used as consistently across schools, so were excluded from analysis. Both PAT Reading Comprehension and

PAT Mathematics have online and paper-based versions. There has been a gradual uptake of online assessment over the past decade. Test mode was not included in the main analysis because the 2012 to 2020 field records show historical results entered the system rather than whether testing was online or on paper. From 2021, online administration can be identified directly, while manually entered records provide only an approximate paper proxy. Refreshed series estimates changed little when mode was included.

For a variety of reasons, students occasionally take more than one PAT assessment in a term, so only the first record per student was retained for analysis each year.

A small minority of refreshed assessment series records from the limited 2023 Mathematics and 2024 Reading Comprehension trials were excluded from the analysis. Legacy records in those years remained in the legacy segment.

TABLE 3. **Primary model-frame coverage by assessment series**

Assessment series	Analysed years	Baseline ¹	Assessments	Unique students	Schools
PAT Mathematics, legacy	2012 to 2023	2012	995,033	448,168	972
PAT Mathematics, refreshed	2024 to 2026	2024	307,813	199,758	1,004
PAT Reading Comprehension, legacy	2012 to 2024	2012	950,679	434,889	936
PAT Reading Comprehension, refreshed	2025 to 2026	2025	246,028	176,566	952
PAT Reading Vocabulary	2012 to 2026	2012	559,019	287,483	754

Records were not excluded solely because either the 2011 or 2017 school-decile snapshot was missing. Those historical decile fields were retained as contextual and potential sensitivity variables, but they did not define the primary analysis population. The primary models instead used simulated EQI, which was available for all records brought into scope by removing the historical-decile exclusion. Table 3 reports the model-complete frames after all other eligibility, deduplication, and complete-case filters.

Assessment refreshes, linking, and visual alignment

The Mathematics and Reading Comprehension refreshes were not unlinked replacements. NZCER psychometrically linked the refreshed material to the established subject scales using Rasch modelling. The refreshed Mathematics items were calibrated onto the existing PAT Mathematics scale, and NZCER describes old and refreshed scale scores as broadly comparable after allowing for measurement error (Darr, 2024). The refreshed Reading Comprehension tests were initially linked to the existing PAT scale through mixed forms and common linking items (New Zealand Council for Educational Research, 2024a).

The amount of change differed between the two assessments. Approximately half of the Mathematics questions were entirely unchanged, and many others were altered only lightly. The refreshed Mathematics assessment is therefore close to the earlier assessment in content. Reading Comprehension changed more substantially, with many new or updated items. NZCER refined the

¹ Baseline refers to the principal change models; heatmap and distributional reference periods are described below.

Reading Comprehension calibrations using 2024 data and then applied a further calibration update to results from 2025 onward in March 2026 (NZCER Assessment Services, 2025, 2026).

The official Rasch analyses address the psychometric question of how assessment scores are to be located on the scale for each subject. However, this report also had to consider a more limited question requiring greater precision. Could the transition from legacy to refreshed assessment series be estimated precisely enough to distinguish it from the achievement score trajectory this study seeks to measure? Exploratory bridge estimates were examined to address this question.

On balance, across the bridging model candidates, the general signal was a small Mathematics improvement between 2023 and 2024, and a small Reading Comprehension decline between 2024 and 2025. However, these directions are uncertain as the error around the bridge estimates is of the same order as the year-to-year differences the modelled series is intended to track.

This uncertainty is more consequential than it might otherwise be because, as discussed in the Introduction, the assessment boundaries coincide with the rapidly changing policy landscape. Transition between assessment series and recent policy changes is therefore potentially statistically confounded. There is a risk that statistical uncertainty in the bridging could remove part of a real change that happened at the same time or amplify a real assessment score movement that could potentially be interpreted as policy related.

In selected plots that display the legacy and refreshed series together, the first full refreshed point is shifted vertically so that it meets the final legacy point. The primary analysis in this report does not apply a numerical bridge. Mathematics and Reading Comprehension are fitted as separate legacy and refreshed segments. For display, the first full refreshed point is shifted vertically so that it meets the final legacy point. This visual alignment is intended as a deliberately conservative reporting convention. It should not be taken as an assertion that actual student achievement was unchanged across the boundary, just that we cannot reconstruct the magnitude of any change with sufficient precision.

Reading Vocabulary was not refreshed and remains one continuous series from 2012 to 2026.

School equity measures

During the period of this study, more than one equity measure was in use by the Ministry. Historically, school deciles were widely used in funding, public reporting, and educational discussion. Deciles grouped schools into 10 bands using census-derived socioeconomic factors, and they were systematically revised after censuses. Deciles were also updated between censuses when a school requested a review (Ministry of Education, 2020). The relatively coarse nature of the decile measure, together with its infrequent and inconsistent revision history, limits its ability to capture changing barriers to achievement over time.

From 2023, deciles were replaced for funding purposes by EQI. The EQI is designed to estimate the extent to which students at a school face socioeconomic barriers to achievement, and the Ministry uses it in annual funding and reporting contexts (Ministry of Education, 2020). The official EQI and the historical decile measure are not directly comparable; they are different measures built for different purposes and with different underlying information structures. That creates a methodological challenge for a longitudinal report that begins in 2012 but wants a single school-level deprivation covariate carried consistently through the whole series.

The solution was to construct a simulated EQI using historical Ministry directory files archived by

NZCER. A monotone-decile gradient boosting model was trained on schools with observed EQI in 2024 to 2026 (in-sample $R^2 = 0.94$) and used to simulate EQI for 2011 to 2026. Predictors included historical decile, school roll, ethnic-composition shares, school authority, school type, urban/rural status, Māori electorate, and region. Historical deciles were smoothed over time using linear interpolation between observed decile anchors to avoid sudden jumps where decile was updated infrequently. Where Ministry categories changed over time, the older, simpler classification was mapped forward where necessary to preserve analytic comparability.

Historical decile and observed EQI fields were retained as contextual variables and for possible alternative-adjustment checks, but they were not used to restrict the primary population. All results presented in this report use simulated EQI as the school-level socioeconomic adjustment.

EQI and simulated EQI are school-level contextual measures. Consequently, these analyses do not estimate socioeconomic differences between individual students or households.

Ethnicity presentation

Three ethnicity coding approaches were explored. The first mapped recorded ethnicities to the Ministry's Level 1 ethnic group—European/Pākehā, Māori, Pacific Peoples, Asian, Middle Eastern, Latin American, and African (MELAA), and Other Ethnicity. The second retained all ethnicities individually without combining them into categories. The third used a Level-1-plus approach, retaining the Level 1 groups for Māori, Pacific Peoples, Other Ethnicity, and European/Pākehā, while separating Asian into North Asian and South and Southeast Asian groupings and separating MELAA into its component Middle Eastern, Latin American, and African groupings. Top-line models were fitted under each coding approach and compared using the Bayesian Information Criterion. The Level 1 approach performed substantially less well than the other two approaches. Although the best performing coding scheme (of the other two) differed by assessment series, the Level-1-plus approach was used throughout the main analyses because it retained important additional detail while remaining computationally feasible for models that included interaction terms.

In all models, Ethnicity was treated as total-response rather than prioritised. In other words, students with more than one recorded ethnicity contributed to each ethnicity category they identified with.

Model specification and statistical estimation

The primary analysis uses the all-students analysis population shown in Table 3. Models were fitted separately to each assessment segment shown there. Fixed effects included assessment year and student year level as categorical factors, gender, centred simulated EQI, and the total-response ethnicity indicators described above. Random intercepts were included for students and schools. The main all-students models include assessment year, student year level, gender, centred simulated EQI, and the retained ethnicity indicators described above. Because ethnicity was modelled as total response, these indicators were allowed to overlap rather than treated as mutually exclusive categories. Random intercepts for student and school were included to account for repeated observations of the same learner and clustering within schools.

The main model family can be written as:

scaleScore ~ Year + Level + gender + simulated EQI + ethnicity indicators + (1 | student) + (1 | school)

This means that the model estimates PAT scale scores as a function of when the assessment was

taken, the student's year level, gender, school socioeconomic context, and ethnicities. The assessment year term is the main trend term; within each modelled series or segment, it estimates how scores in each year differ from the relevant within-series baseline, after accounting for the other variables in the model. Student year level adjusts for the fact that older students are expected to have higher scale scores than younger students. Gender, simulated EQI, and the ethnicity indicators adjust for differences in the composition of the students and schools contributing data in each year. Because ethnicity is treated as total-response, students with more than one recorded ethnicity contribute to all relevant ethnicity indicators. The student random intercept accounts for repeated observations of the same learner, while the school random intercept accounts for similarity among observations from the same school.

Additional analyses extend this structure by adding the interaction needed for further exploration of additional questions. For example, the year-level chapter uses an interaction term between calendar year and student year level. The gender chapter uses an interaction between calendar year and gender. The ethnicity chapter uses an interaction between calendar year and ethnic group. The regional chapter uses an interaction between calendar year and region. In the Mathematics and Reading Comprehension subgroup models, calendar year by year level controls are also retained so that level-specific movement is not absorbed into subgroup trajectories around the refresh.

Modelled results are reported as scale-score estimates with 95% confidence intervals. In legacy change plots, the plotted estimate is the modelled difference between the relevant assessment year and 2012. In refreshed change plots, the estimate is the modelled difference from 2024 for Mathematics or 2025 for Reading Comprehension. Selected Mathematics and Reading Comprehension plots then apply the visual alignment described above. On single-series figures, the left axis retains the legacy change from 2012 and the right axis reports refreshed change from the first full refreshed year. In subgroup figures, each line is aligned separately so that a common shift does not create artificial subgroup discontinuities. For heatmaps without displayed confidence intervals, legacy Mathematics, legacy Reading Comprehension, and Reading Vocabulary use equal-calendar-year 2012 to 2014 references, while refreshed panels use the first full refreshed year. Model intervals are shifted by the same display offset as the point estimate. They describe ordinary within-segment model uncertainty and do not include uncertainty for a cross-boundary bridge that was not applied.

Predicted annual estimates are standardised to a specified reference composition rather than calculated using the observed covariate mix in each assessment year. For Mathematics and Reading Comprehension, both segments are standardised to the 2026 refreshed-series reference population. This gives the two segments a common modelled composition without imposing a numerical score bridge. Reading Vocabulary also uses a 2026 reference population. In practical terms, this means that each year's average was estimated as if the students tested had the same mix of year levels, genders, ethnicities, and school EQI values as the students tested in that subject in 2026. Using the same 2026 mix for every year reduces the risk that changes in who took PAT are mistaken for changes in achievement. The estimates describe the average for this common mix, rather than allowing changes in who sat PAT assessments each year to affect the results presented.

Distributional analysis

Adjusted quantile change was estimated by fitting distribution regression models to the proportion of students at or below each observed PAT score in each assessment year and year level. Predictions were standardised to the same empirical mix of recorded gender, total response ethnicity, and simulated school EQI, then combined across Years 4 to 8 using fixed weights. The adjusted cumulative

distribution was converted to a score at each percentile rank. Change is the difference between an annual quantile and the corresponding within-series baseline quantile. For example, -2 at P25 means that the adjusted 25th percentile was two PAT scale-score points below its baseline.

Results were averaged across lower P10 to P30, middle P30 to P70, and upper P70 to P90 rank bands. Legacy Mathematics, legacy Reading Comprehension, and Reading Vocabulary used an equal calendar year 2012 to 2014 baseline. Refreshed Mathematics used 2024 and refreshed Reading Comprehension used 2025. Joint maximum simultaneous 95% intervals were used for claims that estimates or rank-band contrasts differed from zero. A continuous P5 to P95 sensitivity analysis checked that conclusions did not depend on the reporting bands. No comparison was made across a refresh boundary.

Year 4 entry and transition analysis

The Year 4 entry and transition analysis used the same all-students model frames and adjustment strategy as the main analyses. Its primary figures follow nominal groups diagonally across repeated annual year-level estimates. Consequently, the Year 4 group observed in one year is treated as the nominal Year 5 cohort in the next, although it need not contain the same students. These figures therefore describe population-level movement and do not require the same students to have been observed in successive years. All numerical comparisons were made within assessment series.

The increasingly favourable Year 4 to Year 5 transition raised the possibility that weaker Reading Comprehension cohorts were making up a small amount of ground. This could not be tested by relating Year 4 scores directly to gains because measurement error can make low starting scores appear to be followed by larger gains even when no genuine convergence has occurred. To account for this issue, a linked analysis followed 108,159 learners who remained in the same school from Year 4 to Year 5, forming 2,292 cohorts in 370 schools. Within each of 20 balanced splits, one group of learners measured the cohort's Year 4 position and the other supplied Year 5 outcomes adjusted for their own Year 4 scores, demographic characteristics, assessment timing, and administration mode. The model compared cohorts within the same school and assessment series, weighted cohorts equally, and clustered uncertainty by school. Further models tested change over time, compared the association across all adjacent transitions through Year 8, and examined whether it was stronger in schools with a consistent same-subject PAT history.

Sensitivity analyses and limitations

The main analyses use the full "all-students" model frame for each series because it retains the largest usable population and gives the most stable annual estimates. The main sensitivity framework comprised four checks designed to assess whether the principal annual patterns changed under narrower or more flexible analysis population definitions.

The "stable students" check included learners with PAT records in at least two model-frame years. The "stable schools" check retained schools with at least five eligible Term 1 records in at least 80% of the years in the relevant segment, with the required number of years rounded down and a minimum of 1 year. The separate "schools tested every year" check retained schools with eligible records in every year from 2020 to 2026. The "school-specific trends" check retained the full all-students frame and extended the school random-intercept term to include school-specific slopes for years since the segment baseline, allowing schools to differ in both their starting level and rate of change over time. Additional targeted checks concerning test mode, score rounding, and linked students are described in the relevant subsections. Together, these checks assess whether the national trends depend on

student recurrence, changing school participation, or a common school-level time pattern.

These checks do not replace the main model. They are used to test how sensitive the findings are to different reasonable ways of defining the analysis frame. If a pattern appears across several model frames, it can be read with more confidence. If the direction or size changes across frames, this may change the interpretation of the primary model results.

The analyses in this report describe trends in PAT data as these assessments were used by schools. They are not based on a purpose-built national monitoring sample. The number of assessments is large, and the data cover many schools and students, but students who sit PAT may differ from students who do not.

The models presented in this report cannot adjust for unobserved changes in attendance, teaching, curriculum, assessment practice, school participation, home learning, digital access, whānau circumstances, local disruption, or migration history. Such factors may contribute to residual variation and may confound an estimated association if they are related to both an included predictor and achievement. Subgroup estimates should therefore not be interpreted causally.

Simulated EQI is also a limitation. It provides a consistent school socioeconomic measure across the full study period, including years before the official EQI existed. It is useful for adjustment, but it is still a derived measure, not an official historical EQI series.

The timing of the assessment data affects interpretation. The report uses Term 1 assessments, so each year's results mainly reflect learning conditions from the previous school year. This is especially important when interpreting the early 2020s. Term 1 results from 2021 and 2022 are closely connected to learning conditions during 2020 and 2021.

This is a study of achievement patterns among students and schools using PAT, rather than a causal evaluation or a probability-based estimate of national achievement. It should be read as complementing, rather than replacing, designed national monitoring.

Software used in this report

All analyses in this report were conducted using R version 4.5.3 or higher. Data wrangling was carried out using the `data.table` package. Mixed effects models were fitted with `lme4`, while `lmerTest` provided fixed effect tests using Satterthwaite approximations for degrees of freedom. Figures were produced with `ggplot2` and base R graphics. NZCER has a ChatGPT Business account which was used for some coding and technical writing support. All analysis output, code, and writing that was AI-assisted was human reviewed.

How to read the graphs

Most change figures use a vertical range at least as wide as average annual progress across Years 4 to 8 in that PAT subject. The minimum ranges are 6 scale-score points in Mathematics, 8 in Reading Comprehension, and 7 in Reading Vocabulary. Wider ranges are sometimes used where needed to show the full estimates and confidence intervals.

The zero line marks the relevant within-series reference. Legacy and refreshed segments should be interpreted within their own series, even where the plots align them visually.

3. Overall trends

The overall trend figures presented in this section establish the broad shape of the three PAT series. The following sections examine year-level, entry and transition, distributional, gender, school socioeconomic context, ethnicity, and regional patterns.

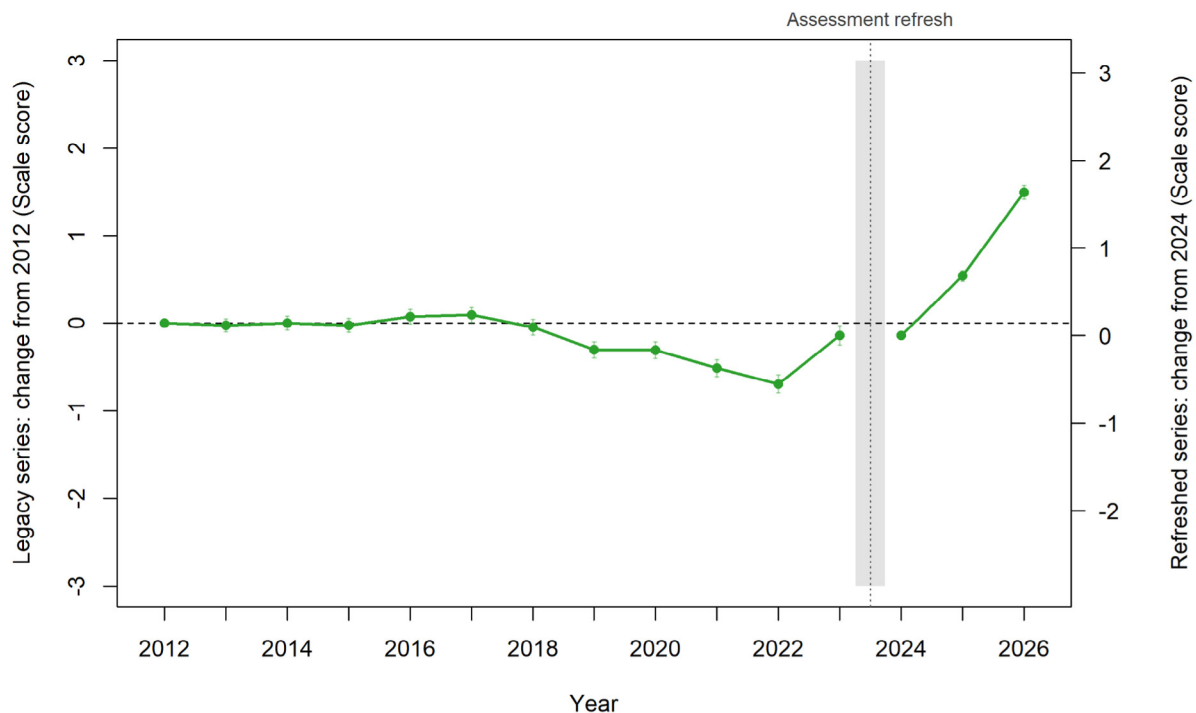


FIGURE 1a. **PAT Mathematics overall national trend with refreshed segment visually aligned**

The Mathematics trend stays close to the 2012 baseline through most of the first half of the legacy series. From 2013 to 2018, the estimates fluctuate modestly around the reference point rather than showing a sustained upward or downward movement. The highest point occurs in 2017, at 0.10 scale-score points above baseline [0.01, 0.18], while the 2018 estimate is close to baseline at -0.05 [-0.14, 0.04].

The pattern changes from 2019. Mathematics is below baseline in both 2019 and 2020, with estimates of -0.30 and -0.30, respectively. The decline then deepens in the early 2020s during the pandemic years, reaching -0.51 in 2021 and -0.69 [-0.79, -0.59] in 2022. The 2022 estimate is the lowest point in the Mathematics legacy series. Given that Term 1 data are used here, each year's data largely reflect learning in the previous year, so the 2021 and 2022 scores correspond closely to the 2020 and 2021 COVID-19 pandemic years (Coblentz, 2026).

By 2023, Mathematics had moved back toward baseline, with an estimate of -0.14 [-0.25, -0.03]. This is a clear improvement from the 2022 trough, but not a full return to the 2012 reference point. Within the refreshed Mathematics series, 2024 is the refreshed baseline. The 2025 estimate is 0.68 [0.62,

0.74], and the 2026 estimate is 1.64 [1.56, 1.71] scale-score points above the 2024 refreshed baseline. The 1.64-point rise is about one-third of average annual progress between Years 4 and 8 on PAT Mathematics. This is strong positive movement within the refreshed series. It is also consistent with the bridge models, which suggested a small positive transition from 2023 to 2024, though that transition-year estimate remains uncertain.

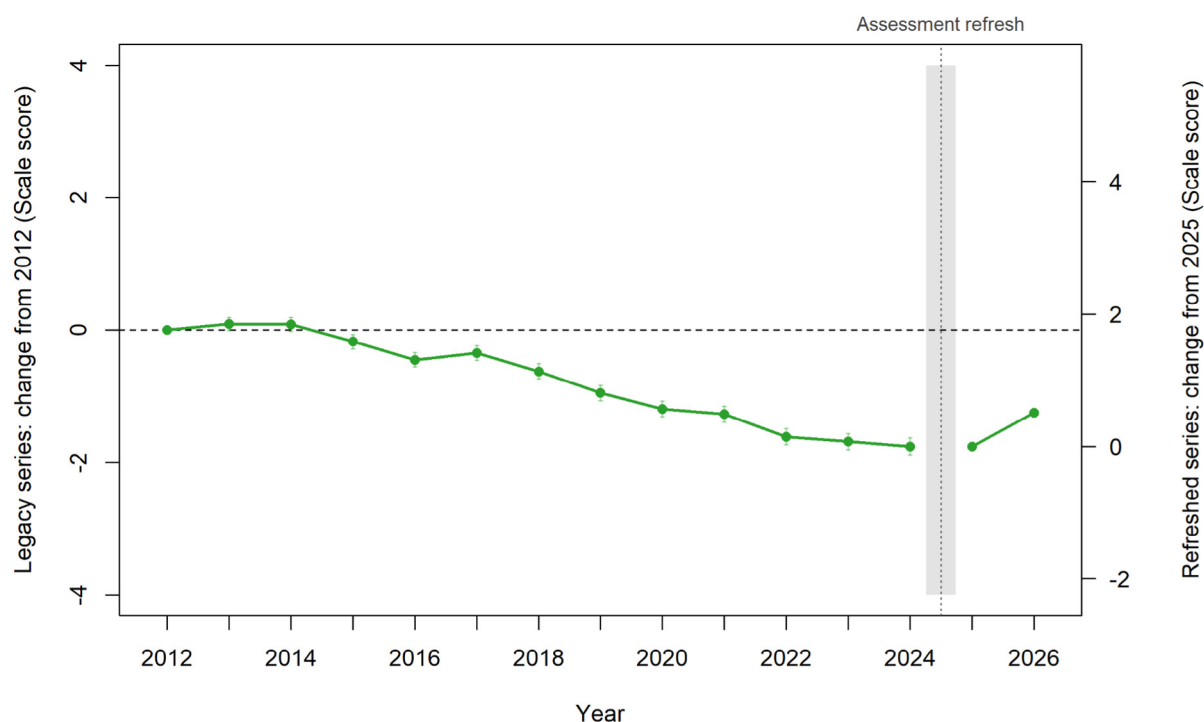


FIGURE 1b. PAT Reading Comprehension overall national trend with refreshed segment visually aligned

Reading Comprehension follows a different path. Unlike Mathematics, the legacy series does not show a late rebound toward baseline. Instead, it moves below baseline by the mid-2010s and remains below baseline for the rest of the legacy period.

The decline is visible before the pandemic years, continues through them, and does not show a legacy-series rebound. The 2024 estimate is -1.76 [-1.89, -1.63], the lowest point in the legacy series. This decline is about one-quarter of average annual progress between Years 4 and 8 on PAT Reading Comprehension. Within the refreshed series, the 2026 estimate is 0.51 [0.45, 0.57] above the 2025 reference, which is less than one-tenth of average annual progress.

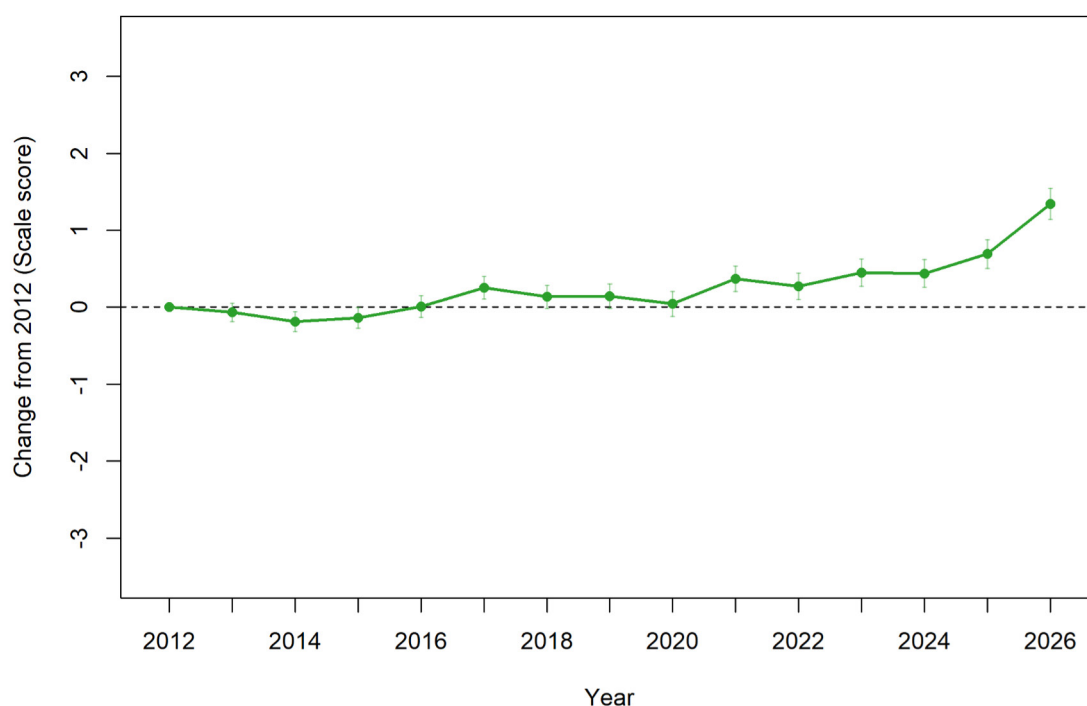


FIGURE 1c. **PAT Reading Vocabulary overall national trend**

Reading Vocabulary does not follow the same pattern as Reading Comprehension. This series is slightly below baseline in some early years, including -0.19 [-0.32, -0.06] in 2014, but then turns positive and remains above baseline through the later period.

From 2017 onward, Reading Vocabulary is generally above the 2012 reference point. The estimate is 0.25 [0.11, 0.40] in 2017, 0.37 [0.21, 0.54] in 2021, 0.45 [0.27, 0.63] in 2023, 0.44 [0.26, 0.62] in 2024, and 0.69 [0.51, 0.88] in 2025. The highest point occurs at the endpoint, with the 2026 estimate reaching 1.34 [1.14, 1.54]. This is about one-fifth of average annual progress between Years 4 and 8 on PAT Reading Vocabulary.

Reading Vocabulary and Reading Comprehension did not move together. Reading Comprehension ended its legacy segment below baseline, while Reading Vocabulary remained on one continuous series and ended above baseline at its highest point.

A matched learner analysis paired Reading Vocabulary and Reading Comprehension records for learners who took both assessments in the same calendar year and at the same year level. It included 464,730 pairs during the 2012 to 2024 legacy Reading Comprehension period and 66,721 during the 2025 to 2026 refreshed period. After giving each year level equal weight, the adjusted Reading Vocabulary score relative to the Reading Comprehension score increased by 2.28 points [1.93, 2.63] from 2012 to 2024. The divergence therefore appeared among learners who took both assessments, although this selected group does not represent all students. No clear further separation was evident from 2025 to 2026, when the estimate was 0.10 [-0.14, 0.34].

Taken together, the three figures establish the main longitudinal frame for the rest of the report. Mathematics shows a pronounced early-2020s dip followed by partial legacy recovery and positive

movement within the refreshed series, with bridge models suggesting a small positive transition at the refresh. Reading Comprehension shows a longer and deeper legacy decline, ending at the most negative point in the legacy series, followed by positive movement from 2025 to 2026 within the refreshed series, while bridge models suggested a small negative transition at the refresh. Reading Vocabulary shows improvement over time concentrated in the second half of the observation window and ends at its highest point in 2026.

4. Year-level patterns

National averages can obscure when changes in assessment scores are occurring in primary education. For example, a pattern concentrated among the younger students tells a different story from one shared across all year levels considered in this report.

Figures 2a to 2c compare each year level with its own within-series baseline. At each series endpoint, a joint Wald test using the estimated covariance matrix assessed whether change relative to each year level's own reference was equal across Years 4 to 8. Pairwise contrasts are reported only when they remained statistically significant after Holm multiple comparison adjustment across the direct year-level comparisons.

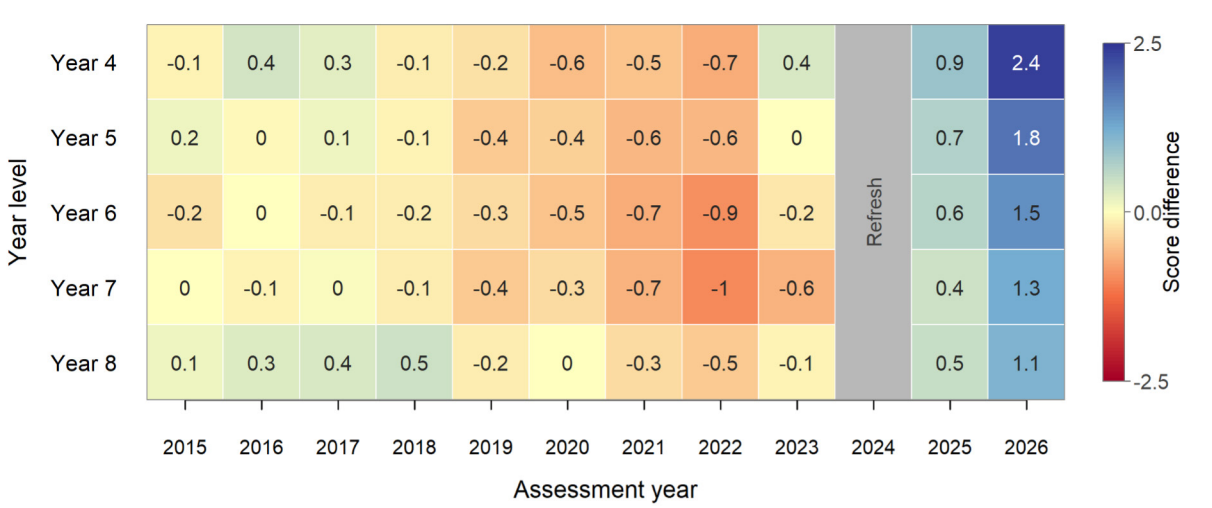


FIGURE 2a. **Adjusted PAT Mathematics score change from the relevant within-series reference, by assessment year and year level**

Mathematics shows an uneven late legacy recovery and a clear age pattern in the refreshed series. Every year level was below its own earlier reference in 2022. By 2023, Year 4 was 0.4 points above its reference, while Year 7 remained 0.6 points below. Direct comparisons confirm that Year 4 was higher than every other level at that endpoint and Year 7 was lower than the others. In the refreshed series, all five levels were higher by 2026. The increase ranged from 2.4 points at Year 4 to 1.1 points at Year 8. The Year 4 increase was clearly larger than the increase at every later level. The other estimates step down with age, although not every adjacent difference is clear enough to treat as statistically distinct.

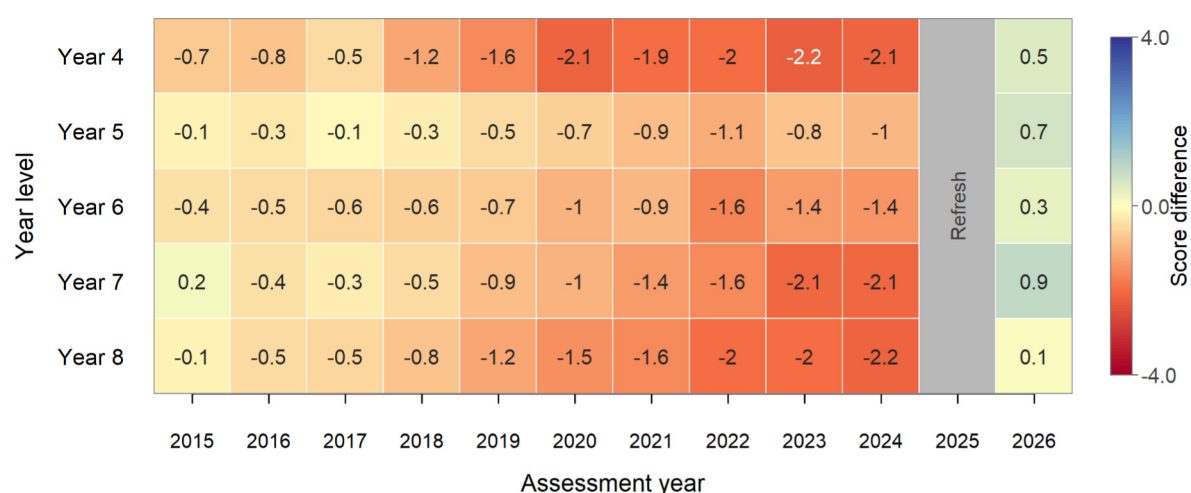


FIGURE 2b. **Adjusted PAT Reading Comprehension score change from the relevant within-series reference, by assessment year and year level**

Reading Comprehension shows broad legacy series weakening rather than a result confined to one age. At the end of the series, Year 5 was least negative at about 1.0 point below, while Year 4, Year 7, and Year 8 were around 2.1 to 2.2 points below. The direct comparison hypothesis testing supports that Year 5 declined less than the other years. The refreshed series rose in point estimate terms at every level from 2025 to 2026. Year 7 had the largest estimate at 0.9 points and Year 8 the smallest at 0.1 points. The direct comparisons confirm that the Year 7 increase was larger than the increases at Years 4, 6, and 8, though it was not statistically distinct from the Year 5 increase of 0.7.

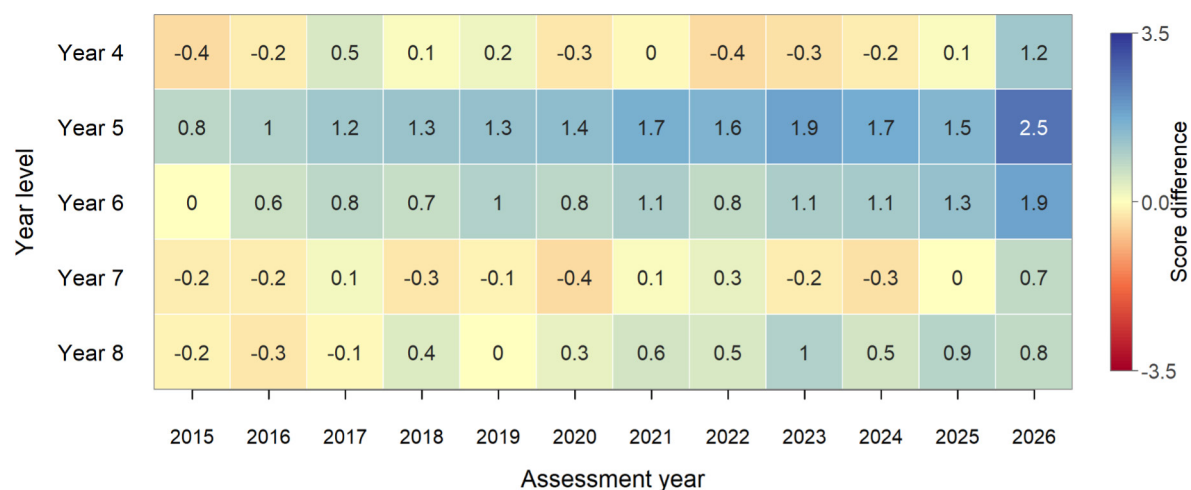


FIGURE 2c. **Adjusted PAT Reading Vocabulary score change from the relevant within-series reference, by assessment year and year level**

Reading Vocabulary ends above its earlier reference at every year level, but the size differs. In 2026, Year 5 was 2.5 points above its reference and Year 6 was 1.9 points above. Year 4 was 1.2 points above, while Year 7 and Year 8 were 0.7 and 0.8 points above. Direct comparison hypothesis testing confirms that Year 5 stands above every other level, and that Year 6 is also stronger than Year 4, Year 7, and Year 8.

The story revealed here differs by subject. The recent Mathematics score increases are larger in the earlier school years. Reading Comprehension legacy series decline was widespread across year levels, with a tentative, early refreshed series improvement that has been weakest at Year 8. Vocabulary strength is greatest in comparison to baseline in Years 5 and 6.

These year-level differences could reflect what successive groups brought into Year 4, what changed later, or both.

5. Year 4 entry and transition analysis

While the previous section explores how the scores of different year levels changed over time, this section instead asks when differences between successive cohorts of students first become visible. It separates differences already present when students enter PAT testing in Year 4 from those that emerge as learners progress through Years 5 to 8. This distinguishes changing achievement at entry from change that accumulates during the primary years. Locating when divergence appears provides a firmer basis for considering how external conditions may be shaping students' learning journeys.

The assessment subjects again show different patterns of change over time. Table 4 summarises the results and the subsequent figures describe these in more detail.

TABLE 4. Year 4 entry and transition summary

Subject	Year 4 entry	Modelled movement through Years 4 to 8	Interpretation
Mathematics	Year 4 scores weakened in the legacy series during 2021 and 2022, following the main COVID-19 disruptions. In the refreshed series, the largest increase in 2026 was at Year 4.	Modelled movement through Years 4 to 7 also weakened in 2021 and 2022, then was positive across all four transitions in 2023. The refreshed transition differences were modestly positive in 2026.	Weakening was visible at Year 4 entry and in modelled movement during the COVID-19 disruptions, followed by a rebound in 2023. The 2026 improvement was strongest at Year 4 but was not confined to entry.
Reading Comprehension	Year 4 scores declined before COVID-19 and remained low through 2024. The refreshed Year 4 entry estimate rose in 2026.	Year 4 to Year 5 movement was usually above its earlier reference, while movement through Years 5 to 7 was less favourable.	Year 4 decline began before COVID-19. The 2026 refreshed results are positive, but only one refreshed interval is available.
Reading Vocabulary	Year 4 remained near the earlier baseline through 2025, then rose in 2026.	Modelled Year 4 to Year 5 movement was strong, Year 6 to Year 7 was weak, and Year 7 to Year 8 was often positive.	The result reflects a pattern across transitions rather than uniform improvement. Relative to each gender's own baseline, the 2026 Year 4 rise was clearer for boys.

5.1 How to read the relative learning figures

Each assessment subject has one heatmap presented below. The top row shows the adjusted Year 4 entry score relative to the relevant Year 4 series reference. The later rows show modelled cohort-level score movement from one year level to the next, compared with movement over the same transition during the reference period. A positive value indicates more favourable score movement than in the reference period. A negative value indicates less favourable score movement.

To assess whether changes in the students tested from year to year affected the findings, the transition models were repeated using only students who could be linked across adjacent assessment

years. The linked student results closely matched the main analysis, with a correlation of $r = 0.98$ and the same direction of change in 95% of comparisons. Differences in direction occurred only when the estimated change was close to zero. This supports the broad timing and direction of the main findings, although linked students are a smaller and more selected group. The main results therefore use all eligible students, with the linked student analysis serving as a sensitivity check.

The markers above the figures provide timing context. The 2021 and 2022 windows reflect learning after the major school closure periods. The 2023 window follows the main disruption period and the initial recovery response. The 2025 assessments are the first completed after a year under the daily hour requirement. The 2026 Mathematics results follow a completed year under the revised 2025 Mathematics content across Years 4 to 8. For Reading Comprehension and Vocabulary, this applies to Years 4 to 6. Revised English content for Years 7 and 8 only became required at the start of 2026. It is important to note that the timing of PAT score changes relative to the external context markers cannot establish causation in the context of this analysis. However, where changes coincide, they may identify questions for further research.

5.2 Mathematics

In the legacy series, both Year 4 entry and movement through several later steps weakened during 2021 and 2022. Recovery was visible in 2023, when all four transition comparisons were positive. In the refreshed series, the 2026 Year 4 entry result was 2.36 points above 2024. The four transition differences were also positive, but much smaller.

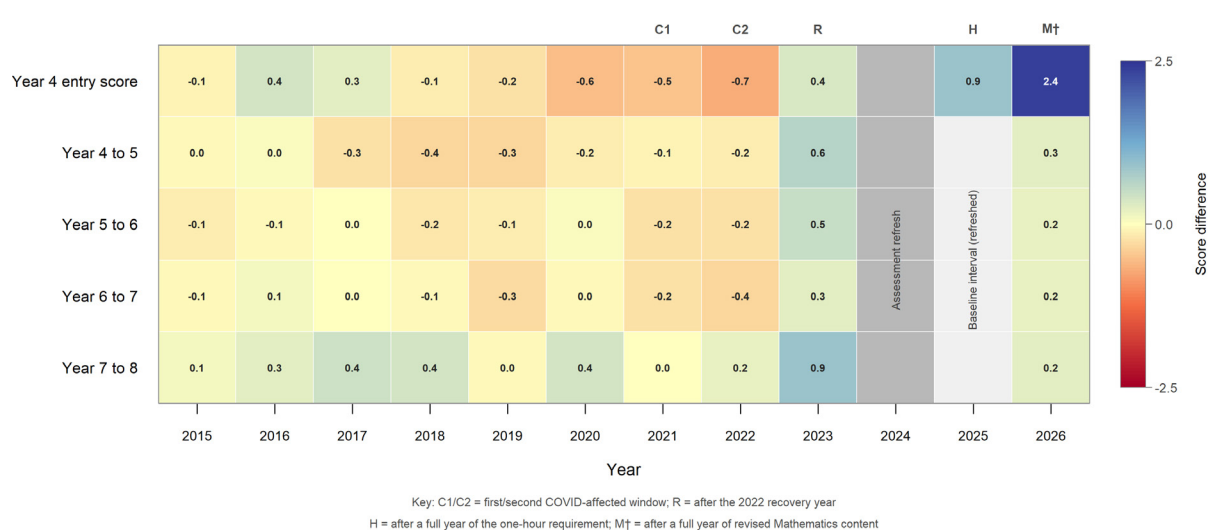


FIGURE 3a. PAT Mathematics Year 4 entry score and later annual interval comparisons for all students

In 2022, movement through Year 4 to Year 5, Year 5 to Year 6, and Year 6 to Year 7 was 0.15, 0.25, and 0.36 points below the earlier reference. Year 7 to Year 8 was 0.18 points above. In 2023, all four were above their references by 0.64, 0.50, 0.27, and 0.89 points. In 2026, the refreshed differences were 0.26, 0.19, 0.18, and 0.22 points compared with the preceding refreshed interval. All four 2026 confidence intervals for the transitions excluded zero. The Year 4 entry difference was substantially larger than any single transition difference. The 2026 pattern was therefore primarily driven by higher scores at Year 4 entry, but not only so.

5.3 Reading Comprehension

Reading Comprehension decline was already visible at Year 4 before COVID 19. The Year 4 estimate was 0.70 points below its earlier reference in 2015, 1.56 points below in 2019, and about 2 points below from 2020 through 2024. Movement from Year 4 to Year 5 was often more favourable, but this was not consistently maintained through the following middle primary steps.

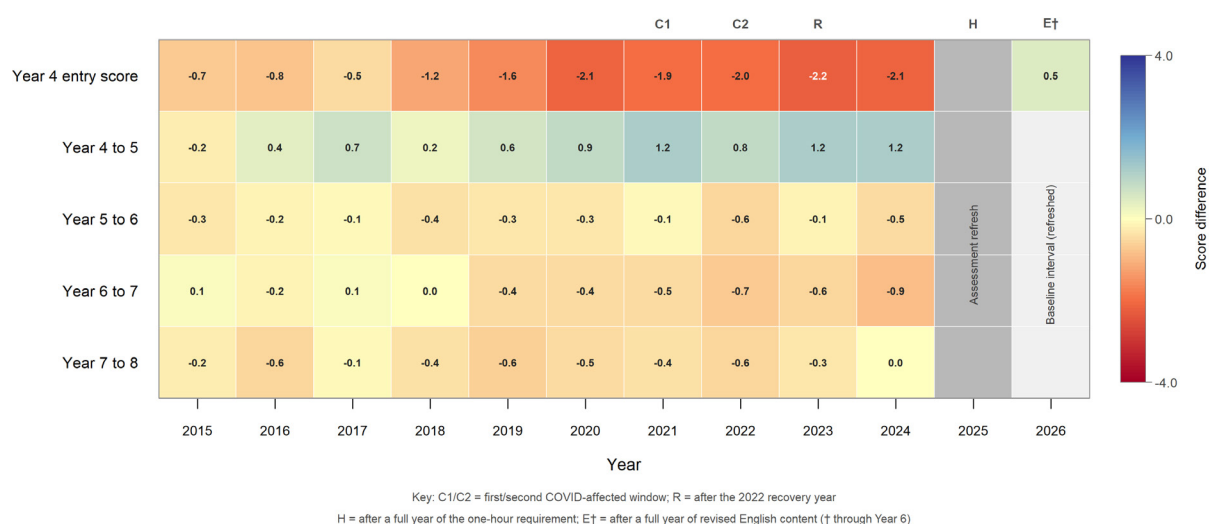


FIGURE 3b. **PAT Reading Comprehension Year 4 entry score and later annual interval comparisons for all students**

Overall, the legacy series results confirm the declining picture for Reading Comprehension seen in earlier sections, with Year 4 scores below the earlier reference and estimated annual learning weaker across several later transitions. Against this broader decline, estimated learning from Year 4 to Year 5 increased over time. It was 0.21 scale-score points below its earlier reference in 2015, moved above the reference in 2016, and, despite some annual variation, reached 0.84 points above in 2022, 1.15 points above in 2023, and 1.22 points above in 2024. This stronger Year 4 to Year 5 learning was not sustained through the later primary school years. In 2024, Year 5 to 6 and Year 6 to 7 remained 0.47 and 0.86 points below their baselines, while Year 7 to Year 8 had recovered from an earlier decline back to its baseline. Greater learning between Year 4 and Year 5 therefore offset some of the weaker starting position at Year 4 but did not fully overcome it or the weaker learning through parts of Years 5 to 7. In the refreshed series, the Year 4 entry estimate and all four annual learning estimates were positive in 2026, as presented in the previous section. Because only one pair of refreshed assessment years is available, it is not yet possible to estimate whether annual learning is strengthening over time within the refreshed series.

An exploratory linked analysis found only a small tendency for unusually weak Year 4 cohorts to perform better than expected in Year 5. The association did not clearly strengthen over time or differ with schools' PAT testing history, so it does not explain the increasingly favourable national transition.

5.4 Reading Vocabulary

Reading Vocabulary shows a recurring transition pattern rather than a uniform rise through every year level. Year 4 entry stayed close to its earlier reference through 2025 and rose in 2026. Movement from Year 4 to Year 5 was consistently strong. Movement from Year 6 to Year 7 was weak, while Year 7 to Year 8 was usually positive.

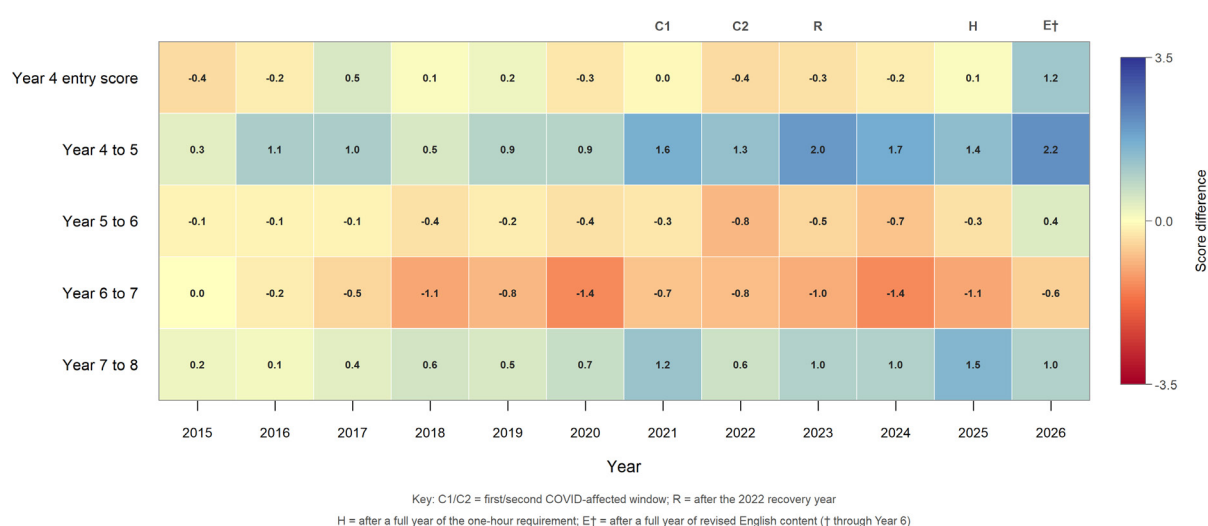


FIGURE 3c. **PAT Reading Vocabulary Year 4 entry score and later annual interval comparisons for all students**

In 2022, the four transition differences were 1.32, -0.83, -0.78, and 0.64 points. In 2024, they were 1.72, -0.72, -1.35, and 0.98 points. In 2026, they were 2.17, 0.43, -0.56, and 0.96 points. Year 5 to Year 6 therefore turned positive in 2026. Year 6 to Year 7 remained below its earlier reference but was less negative than in 2024 and 2025. The 2026 Year 4 rise was clearer for boys relative to their own earlier position, but the later differences were not consistently more favourable for boys.

5.5 What the transition evidence adds

The analysis presented in this section changes how the recent national patterns should be read. In Mathematics, most of the large 2026 improvement is already present when the cohort reaches Year 4. The Year 4 entry estimate is 2.36 points above the 2024 refreshed reference, while the four later transition differences range from 0.18 to 0.26 points. Those later differences are much smaller, but all are positive, with confidence intervals that exclude zero. Gains of this size can accumulate across successive transitions, helping a strong entry cohort retain or extend its advantage as it moves through later years. The national Mathematics increase is therefore led by a much stronger incoming cohort, with smaller gains added later.

Reading Comprehension followed a different path. Successive cohorts entered the PAT window at Year 4 with progressively lower scores. More favourable movement from Year 4 to Year 5 partly offset, but did not reverse, this declining starting position. Less favourable movement through parts of Years 5 to 7 sometimes widened the difference further. The longer decline was therefore driven primarily by falling Year 4 entry scores, with later transitions failing to close the gap. Reading Vocabulary was more uneven. Strong movement from Year 4 to Year 5 and from Year 7 to Year 8 was partly offset by

consistently less favourable movement from Year 6 to Year 7. Its positive endpoint reflects contrasting transition patterns rather than steady improvement across Years 4 to 8.

A linked learner comparison model adjusted for prior PAT score, test interval, assessment year, gender, simulated EQI at the Year 6 origin school, region, urban or rural classification, and school authority. After adjustment, Year 6 to Year 7 progress was lower for learners moving to an intermediate school by 0.50 points [0.28, 0.72] in legacy Mathematics, 0.85 [0.46, 1.24] in refreshed Mathematics, 0.67 [0.38, 0.95] in legacy Reading Comprehension, 0.50 [0.03, 0.98] in refreshed Reading Comprehension, and 0.65 [0.23, 1.07] in Reading Vocabulary. In the available 3-year panels, the two pathways showed no clear difference in Year 5 to Year 6 progress before the transition. In the 4-year panels, the difference improved during Year 7 to Year 8 in legacy Mathematics and Reading Comprehension, reversing direction in Mathematics. This pattern is consistent with a temporary interruption associated with transition, particularly in Mathematics and Reading Comprehension, but does not show that intermediate schools caused it. Entry and transition patterns show when change became visible. They do not show whether it was shared evenly across the score distribution.

6. Changes across the score distribution

Changes in the average score can conceal different patterns across the score distribution. The same average movement could be broadly shared or disproportionately concentrated among learners at only certain relative ability levels. To distinguish these possibilities, adjusted score change was summarised for three broad bands. The lower band covers P10 to P30, the middle band P30 to P70, and the upper band P70 to P90. These are descriptive groupings chosen to make the results readable, not natural divisions between students. A continuous analysis from P5 to P95 was also used to confirm that the conclusions did not depend on the selected boundaries.

The broad quantile band rows presented below average adjusted quantile change within P10 to P30, P30 to P70, and P70 to P90. Values are PAT scale score changes from the relevant within-series baseline after standardising each year to the same empirical mix of recorded gender, total response ethnicity, and simulated school EQI. Lower-middle and upper-middle subtract the middle band. A negative contrast means that the outer band changed less favourably.

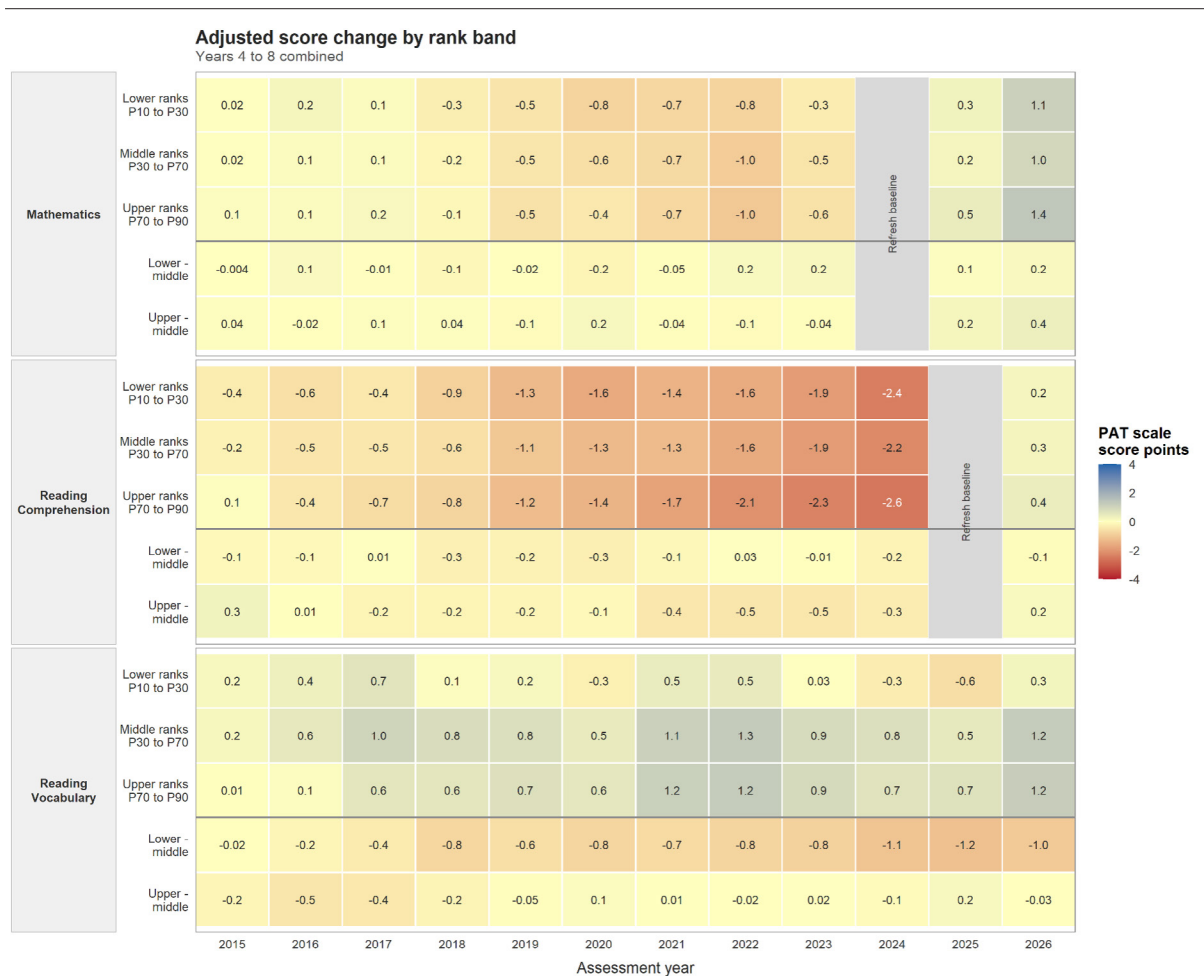


FIGURE 4. Adjusted score change, by rank band, Years 4 to 8 combined

Legacy Mathematics mainly shifted downward rather than changing differently across the distribution. For Years 4 to 8 combined, the average adjusted change across P10 to P90 was below the 2012 to 2014 baseline from 2019 through 2023 and reached -0.94 PAT points in 2022. Neither rank-band contrast showed a sustained difference. In the refreshed series, the corresponding average change was 0.30 points in 2025 and 1.12 in 2026. The 2026 gain was not completely even. The upper band rose 0.38 points more than the middle, while the lower Year 8 band rose 0.63 points less than the middle. The strongest reading is broad recent improvement with some lag in the lower rank band, particularly at Year 8.

Reading Comprehension shows the clearest sustained decline. The combined P10 to P90 average was -0.49 points in 2016 and reached -2.35 by 2024. At Year 4, the 2024 decline was broadly distributed across rank bands. At Year 8, the lower, middle, and upper bands were -1.94, -2.41, and -3.22 points. The upper band fell 0.82 points more than the middle, with a simultaneous interval from -1.30 to -0.33. This pattern was evident from 2017 through 2024. Reading Comprehension therefore fell broadly, with disproportionate loss at the upper part of the Year 8 distribution rather than at the lower part. The refreshed combined P10 to P90 average increased by 0.29 points in 2026, but neither rank-band contrast excluded zero.

Reading Vocabulary presents a different pattern. Lower rank bands changed less favourably than the middle in every year from 2018 through 2026. In 2026, the lower band increased by 0.25 points, while the middle and upper bands increased by 1.24 and 1.22 points. Lower-middle was -0.99 points, with a simultaneous interval from -1.54 to -0.45. The lower part of the combined distribution therefore did not share the clear gains found in the middle and upper parts. This does not show that lower rank bands declined in 2026, and the separate Year 4 and Year 8 contrasts were less precise.

Overall, most score change was broadly shared across the adjusted distribution, though this picture needs some qualification for each assessment subject. The upper Year 8 Reading Comprehension rank band declined more, the lower Reading Vocabulary rank band gained less in the combined distribution, and the lower Year 8 Mathematics rank band shared less in the 2026 refreshed gain. A continuous P5 to P95 location and scale sensitivity analysis supports these broad conclusions. Lower performance may reflect socioeconomic or other forms of disadvantage, and lower rank bands may contain a higher proportion of disadvantaged students. School-level EQI does not capture individual socioeconomic circumstances, so less favourable change in lower rank bands is not direct evidence of widening socioeconomic inequality. The results describe adjusted repeated cross sections rather than causes or individual progress, and PAT score magnitudes are not directly comparable across subjects.

7. Gender patterns

The national, year-level, entry and transition, and distributional analyses show different patterns across the three PAT measures. Next, we turn to the question of whether these patterns differed between girls and boys.

The trajectory figures show how girls' and boys' adjusted scores changed within each assessment series. Table 6 reports the endpoint differences by year level. Each difference is calculated as boys' scores – girls' scores. Positive values mean boys' scores are higher and negative values mean girls' scores are higher.

TABLE 5. Gender change summary, by assessment series

Measure	Years	Girls' change from baseline [95% CI]	Boys' change from baseline [95% CI]	Change in boys–girls difference from baseline [95% CI]	Interpretation
PAT Mathematics	Legacy, 2012 to 2023	-1.33 [-1.47, -1.18]	1.10 [0.95, 1.24]	2.42 [2.24, 2.61]	Boys began with slightly higher scores, and the difference widened because girls' scores fell while boys' scores rose.
	Refreshed, 2024 to 2026	1.20 [1.10, 1.30]	2.07 [1.97, 2.17]	0.87 [0.75, 1.00]	Both groups rose; boys rose more.
PAT Reading Comprehension	Legacy, 2012 to 2024	-2.40 [-2.57, -2.22]	-1.06 [-1.23, -0.88]	1.34 [1.11, 1.56]	Girls began and ended with higher scores, but their advantage narrowed because girls' scores fell more.
	Refreshed, 2025 to 2026	0.40 [0.32, 0.49]	0.58 [0.50, 0.67]	0.18 [0.07, 0.29]	Both groups rose, but boys rose slightly more, narrowing girls' advantage.
PAT Reading Vocabulary	Continuous, 2012 to 2026	0.28 [0.02, 0.54]	2.48 [2.22, 2.74]	2.20 [1.87, 2.52]	Boys' score gains drove the positive overall movement.

Table 5 summarises the main gender result. Across all three measures, boys' scores increased relative to girls' scores within the relevant assessment series. In Mathematics, boys' scores were initially only slightly higher, but the difference grew substantially and widened further within the refreshed segment. In Reading Comprehension, girls' scores remained higher, although the difference narrowed within the legacy series and slightly again within the refreshed segment. In Reading Vocabulary, girls initially had higher scores, but stronger gains among boys produced convergence and, by 2026, slightly higher scores among boys.

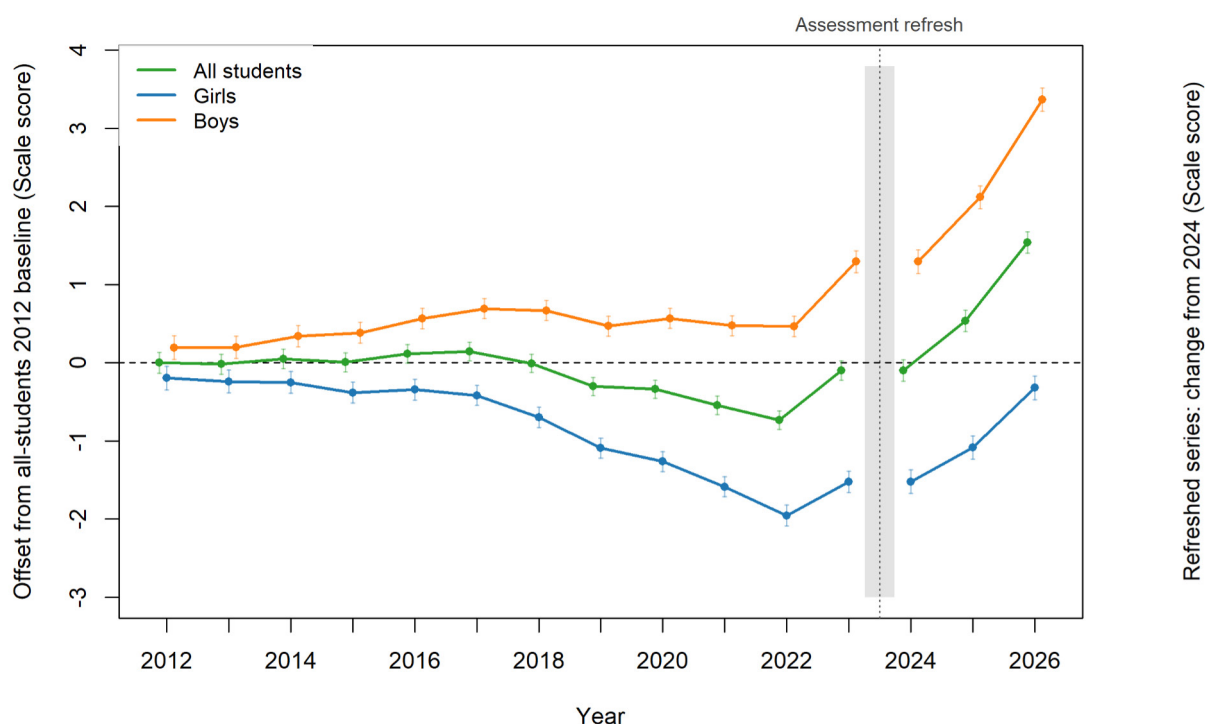


FIGURE 5a. **PAT Mathematics adjusted gender trajectories relative to the all-students baseline**

Mathematics shows the strongest gender divergence. Girls' adjusted trajectory moved from -0.20 relative to the all-students 2012 baseline in 2012 to -1.52 in 2023, corresponding to a within-legacy change of -1.33 scale-score points. Boys moved from 0.19 to 1.29, a change of 1.10 points. The boys-girls difference therefore widened by 2.42 points within the legacy series. This corresponds to approximately half a year level.

The late-series Mathematics dip was not shared evenly. The girls' line was already below baseline before the COVID period, but the 2021 to 2022 decline was largely a decline in girls' scores. Girls moved from -1.27 in 2020 to -1.95 in 2022, while boys remained above their 2012 position, moving from 0.57 to 0.46. In other words, the overall 2022 Mathematics COVID-19 era decline in scores was driven much more by girls' movement than by boys' movement.

The 2023 rebound was also uneven. Girls improved from the 2022 low point but remained well below their 2012 baseline. Boys rose sharply and ended the legacy series at their highest point within it. Within the refreshed segment, both groups rose from 2024 to 2026, but boys again rose more. Girls increased by 1.20 scale-score points and boys by 2.07, widening the boy-girl difference by another 0.87 points.

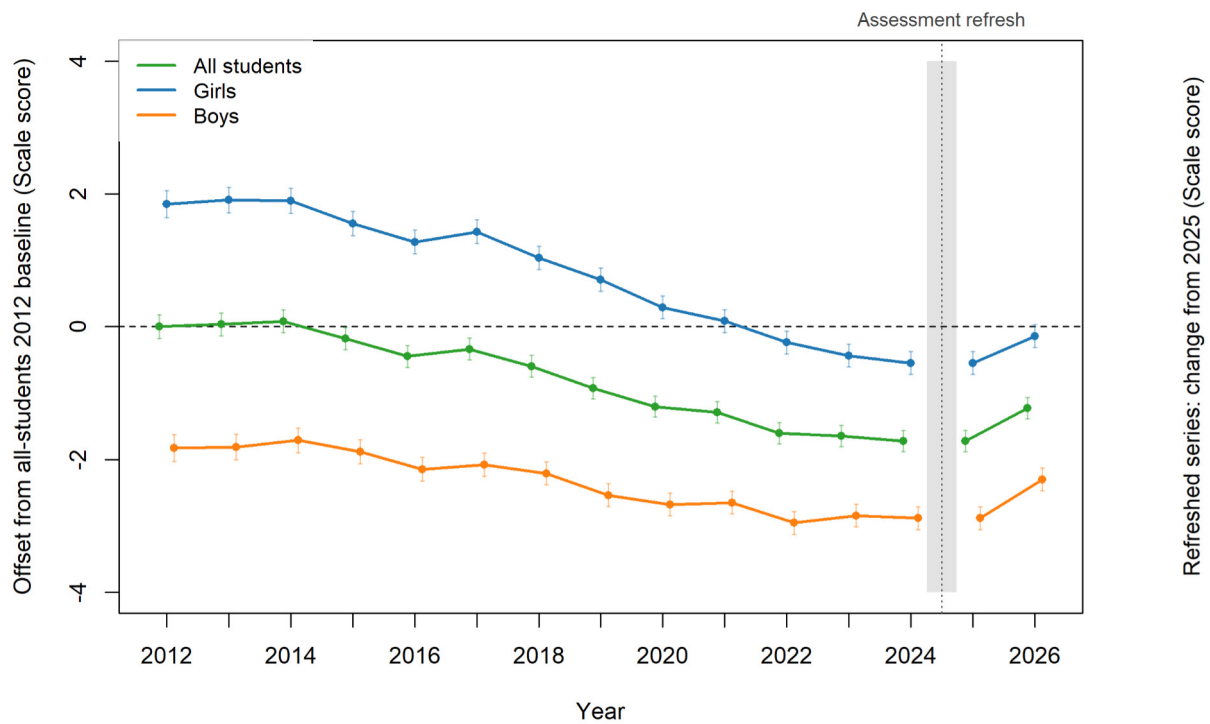


FIGURE 5b. **PAT Reading Comprehension adjusted gender trajectories relative to the all-students baseline**

Reading Comprehension has a different pattern. Girls remained ahead of boys throughout the series, but both groups declined, and girls declined further from their own 2012 starting point.

Girls' adjusted trajectory changed by -2.40 scale-score points from 2012 to 2024, while boys changed by -1.06. The boys–girls difference therefore moved 1.34 points in boys' direction, even though girls were still higher at the end of the legacy series.

The COVID period continues the pre-existing pattern. Girls were already below baseline before 2020, but their decline deepened through the 2021 and 2022 assessment points. Boys also declined, but less steeply. The narrowing of the gender gap should therefore not be read as a simple improvement for boys. It is mainly a story of both groups moving downward, with girls falling further from their 2012 baseline.

Within the refreshed segment, both groups improved from 2025 to 2026. Girls rose by 0.40 [0.32, 0.49] and boys rose by 0.58 [0.50, 0.67]. The difference between girls' and boys' scores therefore narrowed by a further 0.18 points within the refreshed segment [0.07, 0.29].

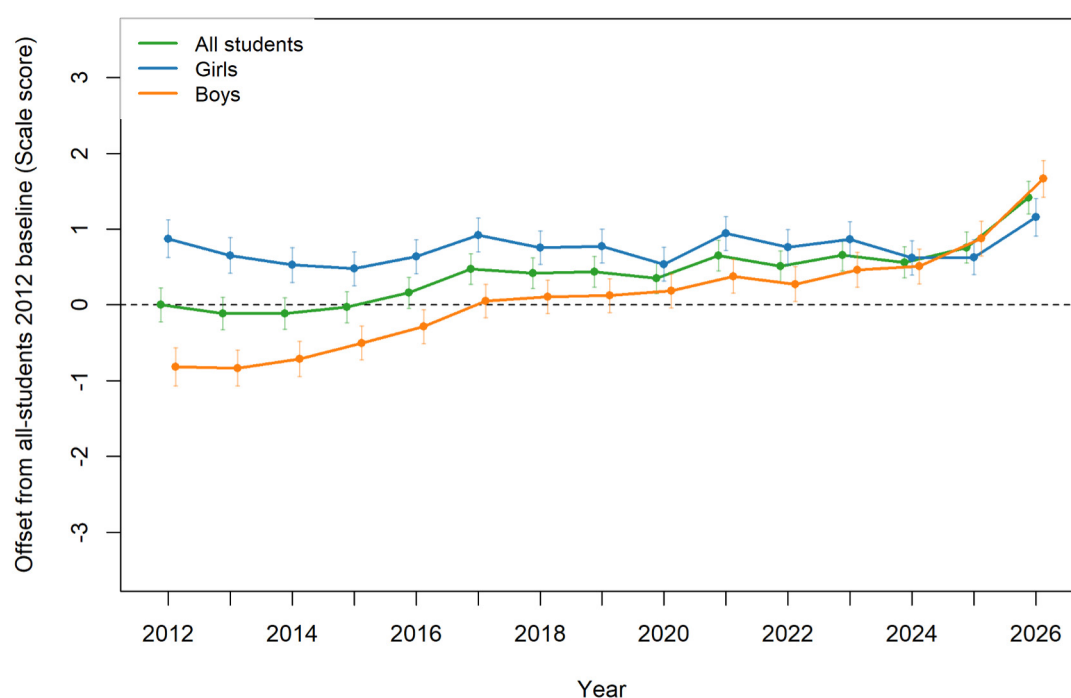


FIGURE 5c. **PAT Reading Vocabulary adjusted gender trajectories relative to the all-students baseline**

Reading Vocabulary shows convergence rather than the same pattern as either Mathematics or Reading Comprehension. Girls moved from 0.87 relative to the all-students baseline in 2012 to 1.16 in 2026, a change of 0.28 points. Boys moved from -0.82 to 1.66, a change of 2.48 points. The boys–girls difference therefore moved 2.20 points in boys’ direction. Overall, the scores converged, with boys scoring slightly higher by 2026.

The Vocabulary pattern around the COVID period points is also unlike Mathematics. There is no equivalent 2021 to 2022 decline. Instead, both groups are higher in 2021 than in 2020, with boys already moving back toward the all-students baseline and girls clearly above it. After 2024, boys continue to rise more strongly than girls, although girls also end above their 2012 position by 2026.

This helps explain why the overall Reading Vocabulary trend is positive. The positive movement is driven mainly by boys’ gains, not by a broad upward shift of the same size for both genders.

TABLE 6. Endpoint boys–girls adjusted gap by year level

Measure and endpoint	Year 4	Year 5	Year 6	Year 7	Year 8
PAT Mathematics legacy, 2023	1.40 [1.05, 1.76]	2.63 [2.30, 2.95]	2.21 [1.92, 2.50]	2.31 [2.04, 2.58]	2.69 [2.42, 2.95]
PAT Mathematics refreshed, 2026	1.68 [1.38, 1.98]	3.27 [3.00, 3.54]	3.20 [2.94, 3.45]	3.15 [2.89, 3.42]	3.49 [3.22, 3.76]
PAT Reading Comprehension legacy, 2024	-3.23 [-3.60, -2.85]	-3.23 [-3.57, -2.89]	-3.12 [-3.45, -2.79]	-2.28 [-2.57, -2.00]	-2.17 [-2.46, -1.89]
PAT Reading Comprehension refreshed, 2026	-3.10 [-3.39, -2.81]	-2.18 [-2.44, -1.91]	-2.84 [-3.10, -2.58]	-2.51 [-2.76, -2.27]	-2.96 [-3.22, -2.70]
PAT Reading Vocabulary, 2026	-0.68 [-1.32, -0.04]	-0.26 [-0.81, 0.29]	1.17 [0.66, 1.69]	-0.76 [-1.30, -0.21]	-0.07 [-0.59, 0.45]

Across all three measures, the boys–girls difference moved in boys' direction. This widened an initial boys' advantage in Mathematics, narrowed but did not remove the girls' advantage in Reading Comprehension, and produced convergence in Reading Vocabulary while year-level Vocabulary gaps remained mixed.

While not presented in detail here, similar gender movement was found across most major ethnicity groups. In legacy Mathematics and Reading Vocabulary, the estimated difference moved towards boys within every retained major group. Legacy Reading Comprehension moved towards boys within European/Pākehā, Māori, Pacific, and South, Southeast, and Other Asian groups, while the movement continued in refreshed Mathematics within five major groups. Changing ethnic composition is therefore a poor explanation for the overall gender shift.

8. School socioeconomic context (simulated EQI)

Simulated EQI is used primarily to account for the large socioeconomic differences between the school communities represented in the PAT data. The EQI scale runs from 344 to 569, with higher values indicating greater socioeconomic barriers. Across the modelled series, the difference between the two ends of the EQI range is associated with a difference in PAT scores of about 10 scale points after accounting for the other variables, which is greater than a year of academic progress on all three assessment subjects considered in this report.

This has implications for the ethnicity results presented next in Section 9. The total observed score difference between ethnicity groups combines differences within similar school contexts with differences arising because groups are unevenly distributed across EQI contexts. Adjusting for EQI removes it from the modelled ethnicity contrasts.

Comparing groups within a common school socioeconomic context helps distinguish differences that remain among students in broadly comparable settings from differences arising because groups attend schools serving very different communities. Without this adjustment, score differences associated with school socioeconomic composition could be attributed to ethnicity. The adjusted comparison therefore shows whether differences remain after this source of variation is taken into account. It does not show the full inequality across the school system, which also reflects how unevenly groups are distributed across school contexts. Greater concentration in schools serving communities with higher socioeconomic barriers is associated with lower observed scores, while greater concentration in more advantaged school communities is associated with higher observed scores. Together, these patterns can make the total observed differences in the real world between groups wider than the adjusted modelled values. The ethnicity results should therefore be read as conditional comparisons. Because EQI is measured at school level, individual socioeconomic disadvantage within schools may still contribute to the remaining differences.

The broad national trajectories changed little when school decile was used instead of simulated EQI, but these alternative adjustments did not show whether the relationship between socioeconomic context and achievement changed over time. To explore this, a separate model was fitted for each assessment series with an interaction between assessment year and simulated EQI. The model predicted scores at two fixed points in the observed EQI distribution. The 20th percentile represented a lower barrier school context and the 80th percentile represented a higher barrier context, with the values ranging from about 383 to 388 and 463 to 472 across the series. Predictions for every year used the same year-level, gender, and ethnic group composition as the first year of that series. Changes in the gap therefore represent changes in the modelled relationship between EQI and scores, rather than changes in demographic composition. The two contexts are modelled comparisons and do not represent the same schools being followed over time.

The adjusted score difference between the lower and higher barrier contexts narrowed by 0.65 points [0.11, 1.20] in legacy Mathematics, 1.18 [0.58, 1.79] in legacy Reading Comprehension, and 1.58 [0.84, 2.32] in Reading Vocabulary. How the narrowing occurred is important. In legacy Mathematics, scores declined in the lower barrier context while remaining broadly stable in the higher barrier context. In legacy Reading Comprehension, both contexts declined, but the lower barrier context declined more. Neither result therefore represents improvement across school contexts. Reading Vocabulary followed a more positive pattern, with scores broadly stable in the lower barrier context and rising in the higher barrier context. Refreshed Mathematics showed no clear change in the gap, at -0.05 points [-0.39, 0.29]. Refreshed Reading Comprehension narrowed by 0.30 points [0.08, 0.52] from 2025 to 2026, although this estimate covers only 1 year of change.

9. Ethnicity patterns

This section explores ethnicity trends over time in PAT data. As discussed earlier in the Methodology section, the analysis reports eight focal groups: Māori; Pacific; European/Pākehā; North Asian; South, Southeast, and Other Asian; African; Middle Eastern; and Latin American.

All ethnicity analyses presented here use total-response ethnicity. This means that a student with more than one recorded ethnicity is counted in every group with which they identify. For example, a student who identified as both Pacific and North Asian contributes to the Pacific total-response group and to the North Asian total-response group. In the Pacific profile, that student is counted on the Pacific side of the comparison, not on the not-Pacific side. In the North Asian profile, the same student is counted on the North Asian side of the comparison. The ethnicity focal groups therefore sometimes overlap since students can identify with more than one ethnicity group.

Total-response groups overlap, and the combinations of recorded ethnicities differ between focal groups. Trajectories therefore describe the named total-response groups rather than mutually exclusive populations. For this reason, change over time for each ethnicity is presented through three views, each of which answers a different question. The overview heatmaps show change for each focal group from its own reference. The trajectory figures show how a focal group and other tested students moved from their own starting points.

The heatmaps show each focal group's change from its own within-series reference; they do not show absolute achievement levels or rank bands. The focal trajectory plots show how the focal populations moved from their own segment starting points. They answer the descriptive question, "How did students identified with this ethnicity move compared with students not identified with it?" The same-starting-mix plots ask a different question, "Did the focal-versus-not-group difference change when both sides were evaluated with the same baseline mix of year level, gender, other recorded ethnicities, and average school socioeconomic context?" If the two views differ, this is evidence that changes in the mix of students within the groups contribute to the descriptive pattern. The analysis cannot determine exactly which changing characteristics account for that difference.

The companion tables report the endpoint estimates used in the figures. The focal and comparison columns describe each trajectory from its own starting point. The final comparison column describes relative change after holding the starting mix constant. The columns answer different questions and should not be subtracted from one another.

Table 7 gives the number of retained assessment records for each focal group in each year and across the full series. N refers to assessment records rather than unique students. Because total-response groups overlap, the group columns should not be added together.

TABLE 7. Retained assessment records, by ethnic group, assessment measure, and year

Year	Māori	Pacific	European/ Pākehā	North Asian	South, Southeast, and Other Asian	African	Middle Eastern	Latin American
Mathematics								
2012	8,807	3,945	39,127	2,582	3,374	659	332	142
2013	10,084	4,596	43,676	3,081	3,867	721	367	160
2014	11,014	4,551	46,150	3,368	4,307	708	424	178
2015	13,581	6,302	52,505	3,860	4,969	740	579	258
2016	15,670	7,961	58,690	4,693	6,526	874	671	375
2017	16,418	8,418	60,286	5,624	7,726	1,027	774	442
2018	18,191	8,848	65,360	5,878	8,038	1,216	794	534
2019	19,822	9,425	65,513	6,247	8,099	1,371	903	616
2020	19,323	9,281	65,104	7,037	8,815	1,787	1,000	704
2021	21,935	11,819	69,798	7,614	10,448	2,217	1,330	806
2022	18,096	9,544	60,400	7,421	10,167	1,994	1,159	797
2023	15,437	8,177	49,264	7,024	8,936	1,825	992	711
2024	19,233	11,621	60,384	9,483	13,595	2,553	1,343	1,040
2025	21,919	13,208	64,072	9,445	16,857	2,766	1,445	1,204
2026	19,900	11,631	57,566	8,994	16,731	2,532	1,341	1,106
Total	249,430	129,327	857,895	92,351	132,455	22,990	13,454	9,073
Reading Comprehension								
2012	6,805	3,450	30,382	1,993	2,751	533	255	107
2013	7,620	4,002	34,802	2,451	3,165	602	291	118
2014	8,322	3,642	37,444	2,780	3,570	574	338	136
2015	10,539	4,909	42,499	3,375	4,173	634	476	216
2016	12,305	6,383	48,446	4,015	5,371	749	552	298
2017	13,030	7,141	50,536	4,736	6,291	861	638	354
2018	13,770	7,549	53,340	4,819	6,269	935	632	415
2019	15,381	8,534	55,752	5,754	7,262	1,197	821	539
2020	15,410	8,691	55,551	6,478	7,873	1,523	920	636
2021	17,558	10,311	59,427	7,683	9,245	1,946	1,146	725
2022	15,880	9,754	56,288	7,955	10,209	1,914	1,185	802
2023	16,612	10,511	57,331	9,621	11,302	2,323	1,314	899
2024	16,228	10,230	54,007	9,383	12,032	2,322	1,229	959
2025	22,470	13,508	71,400	12,778	18,332	3,019	1,590	1,312
2026	22,928	14,364	72,457	13,062	21,524	3,117	1,728	1,442
Total	214,858	122,979	779,662	96,883	129,369	22,249	13,115	8,958
Reading Vocabulary								
2012	4,748	2,631	21,582	1,452	1,974	424	213	82
2013	5,027	2,829	23,587	1,652	2,203	421	230	86
2014	5,800	2,926	25,380	1,877	2,563	425	277	102
2015	6,757	3,165	28,155	2,043	2,828	418	339	161
2016	7,134	3,477	29,741	2,275	3,368	451	324	197
2017	7,614	3,727	30,442	2,477	3,399	506	330	218
2018	7,300	3,849	29,409	2,780	3,383	523	365	250
2019	7,417	4,089	27,651	2,681	3,633	635	425	295
2020	7,029	4,055	25,623	2,825	3,637	710	444	290
2021	7,372	4,751	25,945	3,312	3,921	840	502	313
2022	5,953	3,930	21,144	2,676	3,277	612	388	317
2023	6,226	4,218	20,565	2,646	3,767	771	395	306
2024	6,237	4,522	21,234	3,787	4,624	890	465	422
2025	6,969	4,505	22,990	3,904	5,643	1,019	582	467
2026	6,117	4,090	18,868	2,352	4,810	768	431	372
Total	97,700	56,764	372,316	38,739	53,030	9,413	5,710	3,878

Counts are smallest for the Latin American, Middle Eastern, and African focal groups, especially in Reading Vocabulary. The wider uncertainty around those estimates should be read with these counts in mind.

9.1 Overview heatmaps

The heatmaps provide the first cross-group view before the chapter turns to the detailed profiles. They compare each later year with the relevant within-series reference for the same focal group and measure. For all three measures, the legacy or continuous heatmap cells use an equal-calendar-year 2012–2014 reference. The refreshed Mathematics and Reading Comprehension cells use the first full refreshed year. A positive value means that the group is above its own within-series reference. A negative value means that the group is below that same reference. The values are scale-score differences, not percentages.

The equal-calendar-year 2012–2014 reference used for the legacy Mathematics and Reading Comprehension panels and the continuous Reading Vocabulary panel makes those heatmap baselines more stable, especially for smaller focal groups. The heatmaps should also be read by measure and segment rather than by colour alone, because the colour scales differ across the three figures and the refreshed Mathematics and Reading Comprehension cells use a different within-series anchor from the legacy cells.

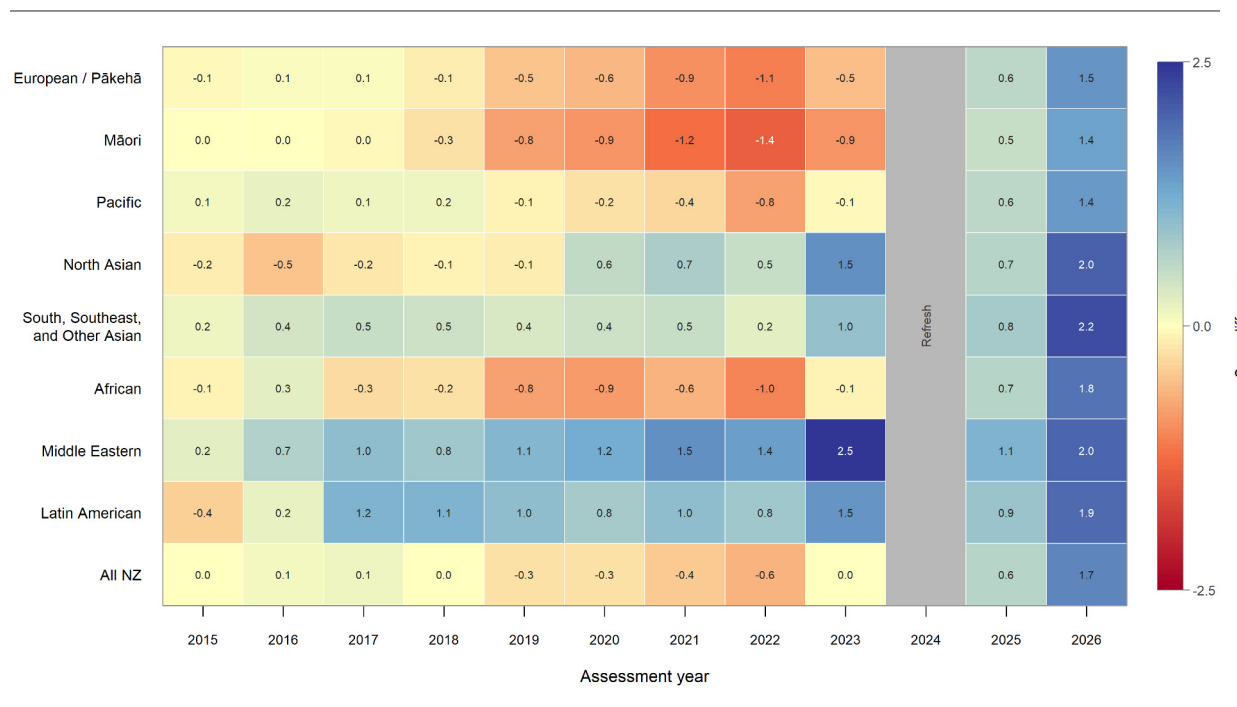


FIGURE 6a. Ethnic group heatmap for PAT Mathematics

The Mathematics heatmap shows spread rather than one shared terminal pattern in the legacy segment and positive movement in the refreshed segment. By 2023, Māori (-0.90) and European/Pākehā (-0.54) are below their 2012–2014 reference. Pacific (-0.06) is close to the reference. North Asian (1.52), South, Southeast, and Other Asian (0.96), Latin American (1.48), and Middle Eastern (2.48) are above the reference. African is slightly below the reference at -0.10.

This fits the earlier Mathematics results, where the legacy segment showed a 2021–2022 dip followed by 2023 recovery, while the refreshed segment rises through 2026. The ethnicity heatmap shows that the legacy endpoint Mathematics pattern is not uniform across focal groups, even though every focal group is above its 2024 refreshed reference by 2026. The 2026 refreshed values range from 1.35 for Māori to 2.20 for South, Southeast, and Other Asian.

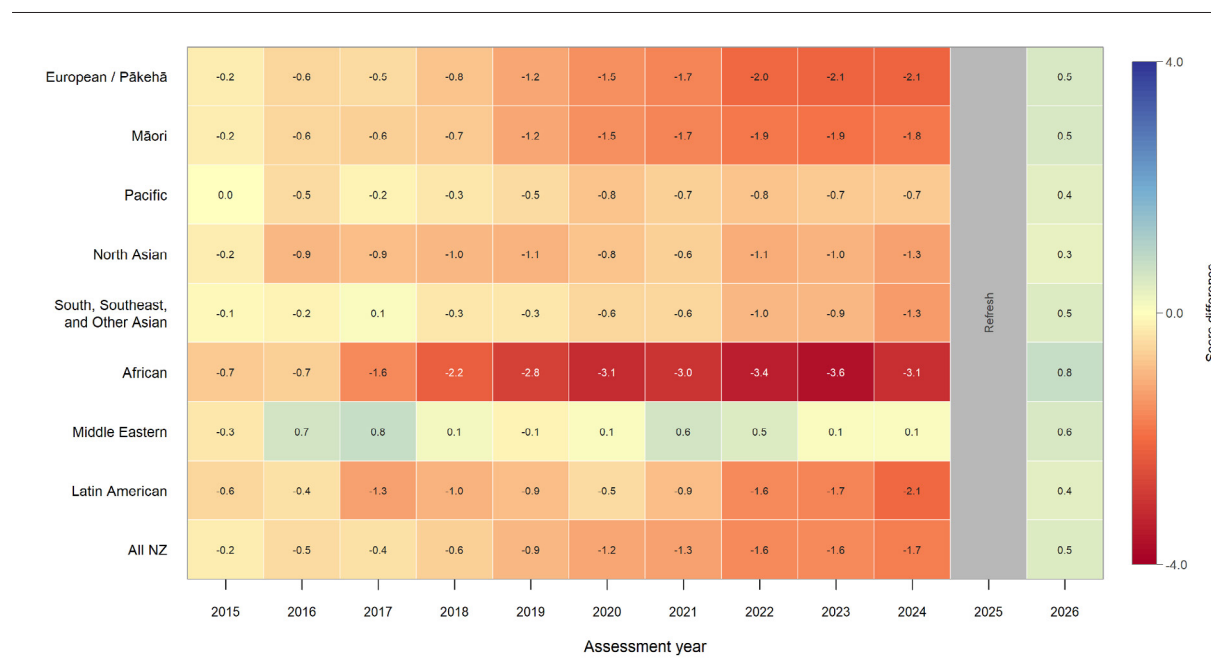


FIGURE 6b. Ethnic group heatmap for PAT Reading Comprehension

Reading Comprehension is much more consistently negative than Mathematics in the legacy segment. By 2024, every focal group except Middle Eastern is below its own 2012–2014 reference. The final legacy values are -1.82 for Māori, -0.72 for Pacific, -2.13 for European/Pākehā, -1.30 for North Asian, -1.34 for South, Southeast, and Other Asian, -3.13 for African, and -2.06 for Latin American. Middle Eastern is near the reference point at 0.05.

The Reading Comprehension pattern has the same negative direction for nearly every group, but its magnitude varies substantially. The 3.18-point spread between those endpoints is equivalent to about two-fifths of average annual progress on the PAT Reading Comprehension scale. The direction is more consistently negative than in the Mathematics or Reading Vocabulary heatmaps. This is consistent with the earlier Reading Comprehension findings across the all-students trend, year-level outputs, and rank band results. Within the refreshed series, each focal group is above its 2025 reference by 2026, with increases ranging from 0.29 for North Asian to 0.76 for African. That positive refreshed movement should be read alongside, rather than as cancelling, the legacy decline.

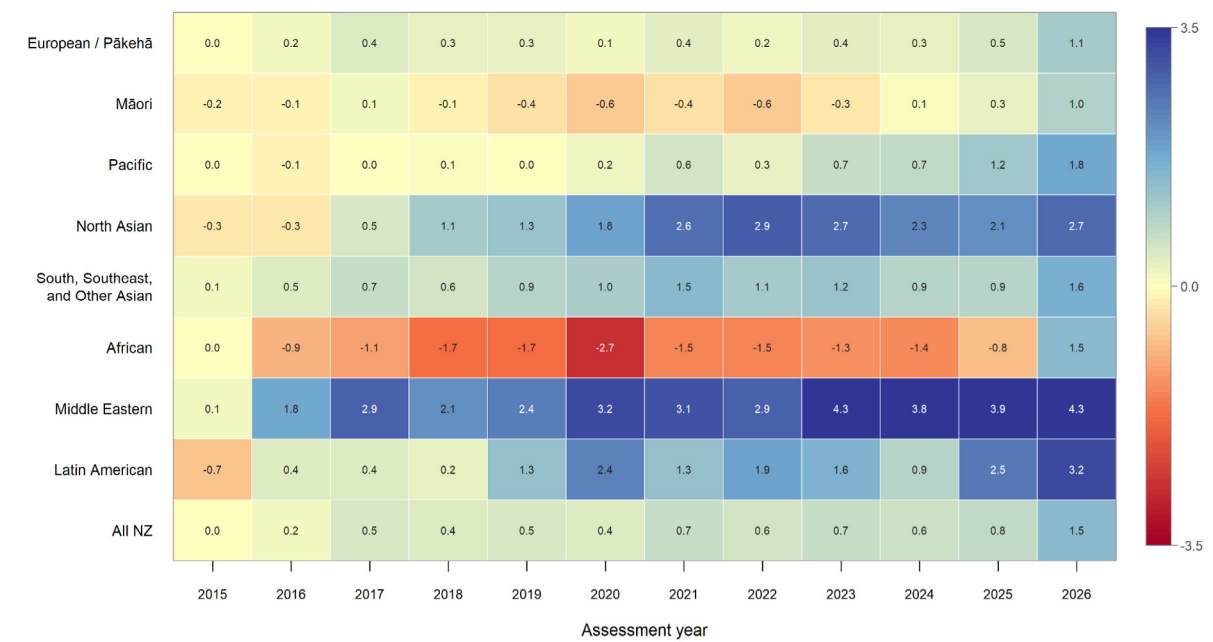


FIGURE 6c. Ethnic group heatmap for PAT Reading Vocabulary

Reading Vocabulary again separates from Reading Comprehension. By 2026, every focal group is above its own equal-calendar-year 2012–2014 reference. Middle Eastern (4.25) has the highest endpoint value, followed by Latin American (3.19) and North Asian (2.68). Pacific (1.82), South, Southeast, and Other Asian (1.65), and African (1.46) are also clearly positive. European/Pākehā (1.12) and Māori (0.97) show more modest positive movement.

The Vocabulary heatmap therefore sits closer to the positive all-students Reading Vocabulary trend than to the Reading Comprehension pattern. At the same time, the smaller-group rows need more caution, especially where annual counts are low.

The heatmap is useful for orientation, but the trajectory profiles below are needed to separate focal movement from relative gap movement.

TABLE 8. Ethnic group overview: terminal-year within-series heatmap values

Ethnic group	Mathematics 2023	Reading Comprehension 2024	Reading Vocabulary 2026	Count and precision note
Māori	-0.90	-1.82	0.97	Large annual counts in all three measures, supporting relatively precise estimates.
Pacific	-0.06	-0.72	1.82	Large annual counts in all three measures, supporting relatively precise estimates.
European/Pākehā	-0.54	-2.13	1.12	Very large annual counts in all three measures, supporting highly precise estimates.
North Asian	1.52	-1.30	2.68	Moderate annual counts in all three measures, supporting reasonably precise estimates.
South, Southeast, and Other Asian	0.96	-1.34	1.65	Moderate annual counts in all three measures, supporting reasonably precise estimates.
African	-0.10	-3.13	1.46	Smaller annual counts, so the estimates are less precise.
Middle Eastern	2.48	0.05	4.25	Smaller annual counts, so the estimates are less precise.
Latin American	1.48	-2.06	3.19	Smallest annual counts, so the estimates are least precise and confidence intervals are wide.

Table 8 gives the terminal values from the three heatmaps. Legacy Mathematics, legacy Reading Comprehension, and continuous Reading Vocabulary are compared with equal-calendar-year 2012–2014 references. Mathematics shows mixed legacy movement, with some focal groups below the reference, some close to it, and some clearly above it. Reading Comprehension is broadly negative, with Middle Eastern the only focal group near its early-series reference. Reading Vocabulary is positive for every focal group by 2026.

Annual counts vary by assessment measure and year. Table 7 provides the counts for each focal result. The smaller focal groups require more caution. This is especially true for the Latin American ethnic group and, in Reading Vocabulary, Middle Eastern.

The heatmaps help summarise the more detailed ethnic profiles that follow. They show where the focal group's own trajectory moved relative to its early-series reference for each assessment series.

9.2 Absolute achievement differences between ethnicity groups

The heatmaps in Section 9.1 show each focal group against its own baseline. This is useful for summarising within-group trajectories, but it does not show the size of the absolute differences between groups. That matters because large achievement differences between social and ethnic

groups are a well-documented feature of the Aotearoa New Zealand education system. The Ministry's review of large education studies describes Aotearoa New Zealand as having one of the largest differences between high- and low-scoring students among OECD or PISA participating education systems, with Pacific learners and ākonga Māori particularly affected (Educational Measurement and Assessment, 2025). NMSSA shows similar primary school patterns. In 2022 mathematics, non-Māori students scored higher than Māori students by 11 scale units at Year 4 and 14 at Year 8, non-Pacific students scored higher than Pacific students by 13 and 17 units, and Asian students scored higher than non-Asian students by 13 and 16 units. NMSSA estimated about eight mathematics scale units as the average progress associated with 1 year of schooling between Years 4 and 8, so differences of 13 to 17 units are roughly equivalent to 1½ to 2 years of average progress on that scale (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2023). In 2019 English reading, Māori and Pacific students were also below the corresponding non-Māori and non-Pacific groups by about 1 year of average progress or more (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2020).

PAT adds to this evidence because it uses high-volume school-use data and uses more detailed total-response ethnicity groupings than are often available. Figures 6d to 6f therefore serve a different purpose from the heatmaps. They place the ethnic groups on the shared PAT scale, showing that the broad pattern of large ethnic group differences reported elsewhere is also visible in PAT data, while also allowing less commonly reported patterns to be seen, especially those for North Asian students. The values in the figures are standardised model predictions from the same model family used elsewhere in the report, with the ethnic group indicator set and the other modelled characteristics held to the common 2026 reference population. For Mathematics and Reading Comprehension, the refreshed segment is visually aligned as in the previous results sections.

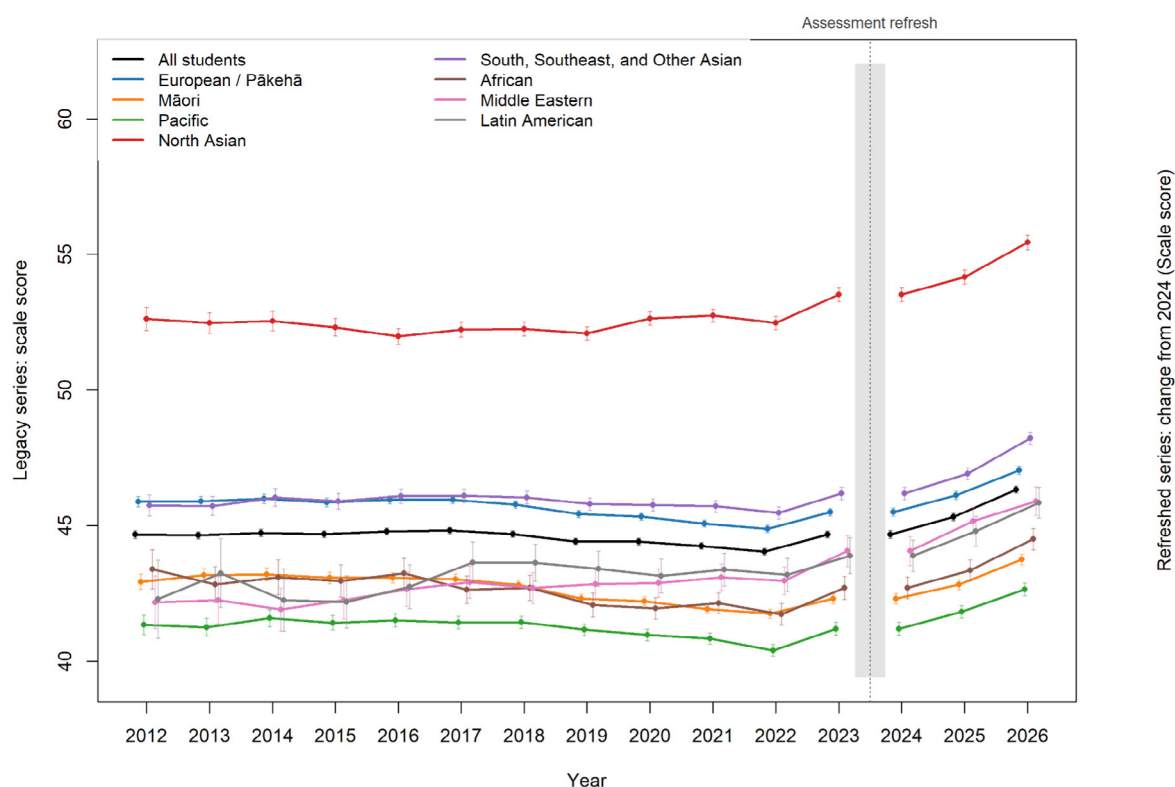


FIGURE 6d. Modelled ethnic group absolute scale-score trajectories for PAT Mathematics

The Mathematics absolute plot shows the strongest separation between groups, and the key contrast is at the top of the distribution rather than only between the top and bottom groups. North Asian students sit well above all other focal groups throughout the period. At the 2026 refreshed endpoint, the modelled North Asian estimate is 55.38 [55.11, 55.64], compared with 48.25 [48.02, 48.47] for South, Southeast, and Other Asian students, the next closest focal group, and 48.06 [47.91, 48.22] for European/Pākehā students. The gap between North Asian students and the next closest group is therefore about 7.1 PAT scale-score points, which is larger than the within-group movements shown in the heatmap and equivalent to more than a year level of academic progress in PAT Mathematics. Pacific, Māori, African, Latin American, and Middle Eastern students sit below the all-students line at the 2026 endpoint. The full top-to-bottom spread is about 11.7 scale-score points, or roughly two school years of progress.

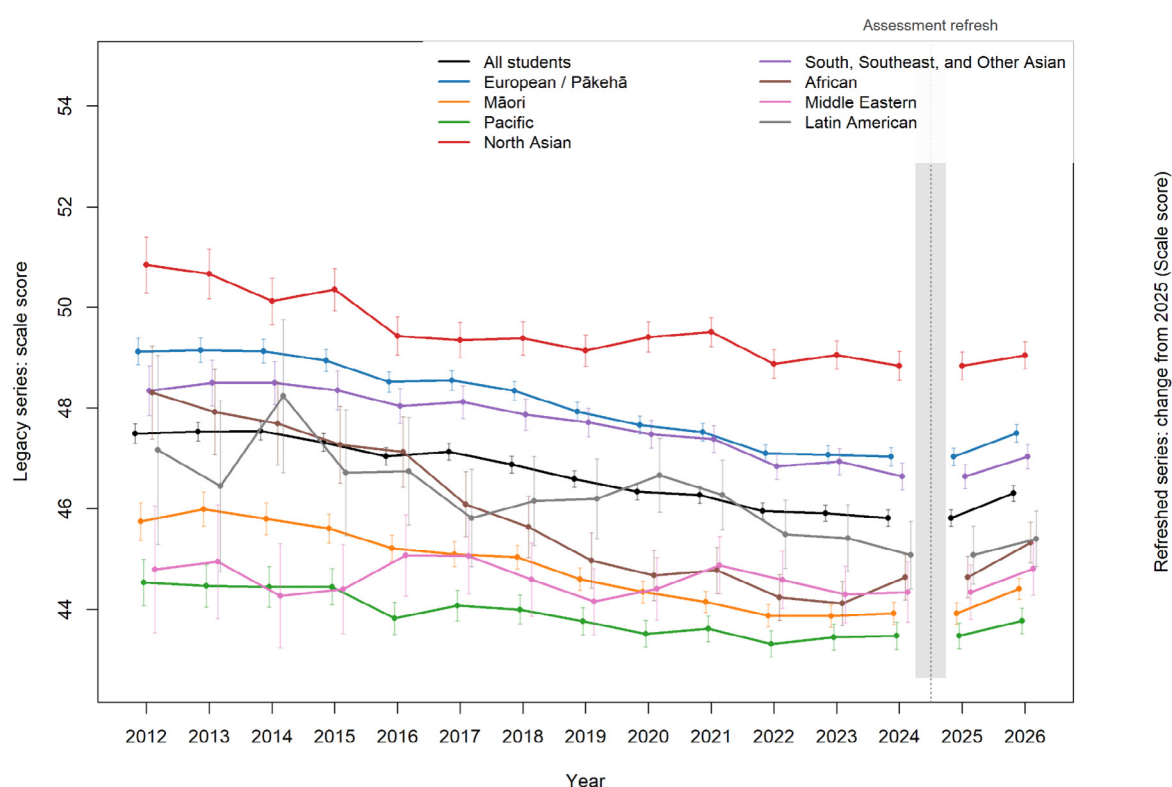


FIGURE 6e. **Modelled ethnic group absolute scale-score trajectories for PAT Reading Comprehension**

The Reading Comprehension absolute trajectory plot shows largely stable separation between groups, but on a smaller scale than Mathematics. The legacy trajectories generally slope downward. At the 2026 refreshed endpoint, the highest modelled estimate is for North Asian students, at 49.99 [49.72, 50.26]. The next highest estimates are European/Pākehā students at 48.89 [48.72, 49.07] and South, Southeast, and Other Asian students at 48.68 [48.45, 48.92]. The distance from the highest group to the next closest group is about 1.1 PAT scale-score points. The full spread is about 4.2 points, from 49.99 to the Pacific estimate of 45.81 [45.56, 46.06]. For context, that full spread is a little over half a year level on this scale.

The Reading Vocabulary absolute plot shows a more compressed version of the same pattern. Vocabulary rises for most focal groups by 2026, and several groups form a close upper cluster rather than one separated line. At the 2026 endpoint, European/Pākehā students have the highest modelled

estimate, at 51.07 [50.81, 51.32], followed closely, and statistically indistinguishably, by African students at 50.70 [49.91, 51.49], South, Southeast, and Other Asian students at 50.53 [50.11, 50.95], North Asian students at 50.34 [49.82, 50.85], and Latin American students at 50.09 [49.00, 51.18]. These estimates are not clearly distinct. The Pacific estimate is 46.65 [46.23, 47.07], giving a total spread of about 4.4 PAT scale-score points, or a little over half a year level.

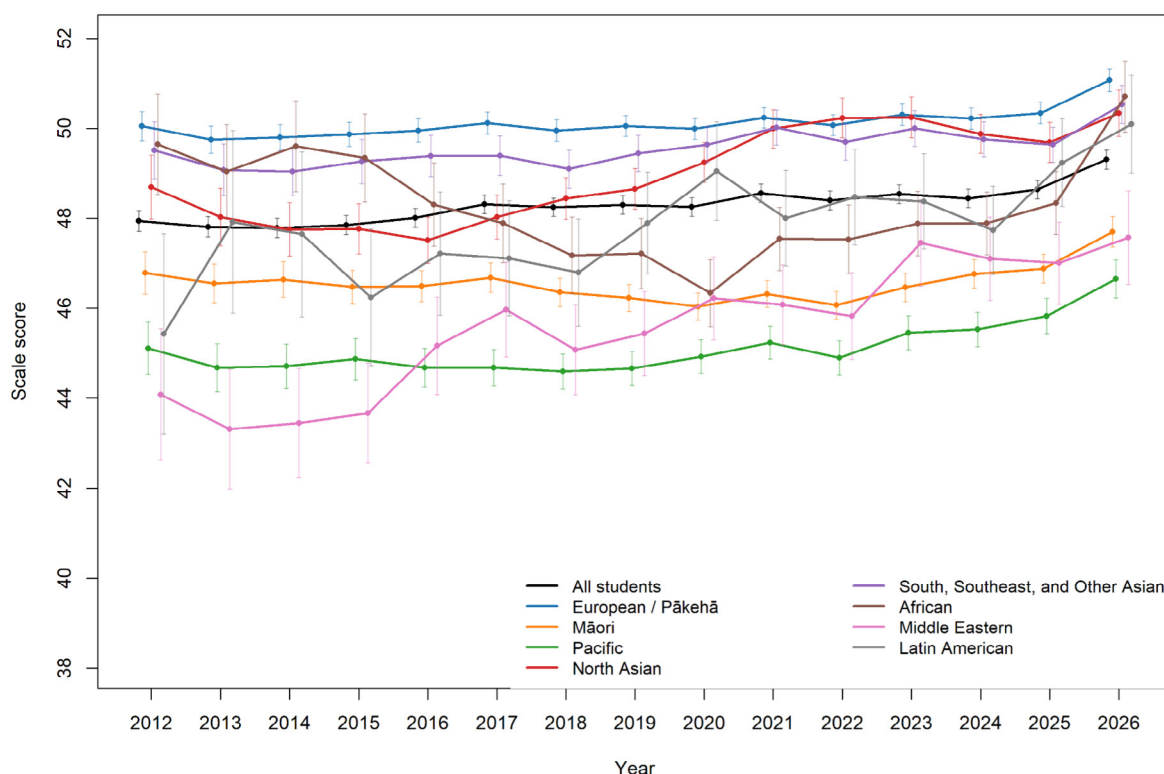


FIGURE 6f. **Modelled ethnic group absolute scale-score trajectories for PAT Reading Vocabulary**

The absolute scale-score trajectory plots shown here confirm the picture described elsewhere that there are wide and stable differences between ethnic groups in Aotearoa New Zealand, while adding nuance that has not been widely discussed because of the more granular ethnicity categories used in this report. In Mathematics, North Asian students are separated from all other focal groups by a large margin, including about 7.1 scale-score points from the next closest group at the 2026 endpoint, while the full focal-group spread is about 11.7 points. Reading Comprehension shows smaller but persistent separation, with a 2026 spread of about 4.2 points and North Asian students again at the upper end. Reading Vocabulary is more compressed and shows a close upper cluster by 2026, but the focal-group spread is still about 4.4 points, or half a year level.

These findings show why absolute levels need to be reported alongside trajectories. Several groups can improve over time while large achievement differences remain, and high-volume PAT data reveal both the familiar national equity pattern and less commonly reported distinctions, especially the North Asian Mathematics profile.

As discussed in Section 8, it is also important to note that these ethnicity trajectories are adjusted for EQI, so actual differences between ethnic groups may be larger or smaller than the modelled differences shown here because the figures hold school socioeconomic context constant rather than allowing the uneven distribution of students across school EQI contexts to contribute to the estimates.

9.3 Māori

This subsection explores how PAT assessment scores for students identified as Māori have changed over time. Table 9 summarises endpoint values. The figures that follow show Mathematics, Reading Comprehension, and Reading Vocabulary trajectories and common-reference difference-over-time outputs. For Mathematics and Reading Comprehension, the plots show the legacy and refreshed assessment segments on the same visual frame, with the refreshed segment aligned for display as in earlier sections.

TABLE 9. Māori companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final non-group trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	-0.63 [-0.93, -0.33]	0.17 [0.04, 0.30]	-0.80 [-1.11, -0.49]
Mathematics refreshed	2024 to 2026	1.45 [1.29, 1.61]	1.70 [1.62, 1.79]	-0.25 [-0.43, -0.08]
Reading Comprehension legacy	2012 to 2024	-1.82 [-2.21, -1.44]	-1.65 [-1.80, -1.50]	-0.18 [-0.57, 0.22]
Reading Comprehension refreshed	2025 to 2026	0.49 [0.34, 0.64]	0.49 [0.42, 0.56]	0.00 [-0.17, 0.16]
Reading Vocabulary	2012 to 2026	0.91 [0.38, 1.45]	1.47 [1.25, 1.70]	-0.56 [-1.12, -0.01]

The focal and non-group estimates come from the trajectory plots. The common-reference gap change estimate shows change in the Māori relative to non-Māori difference from the relevant segment baseline when both sides are evaluated using the same reference population. These columns are not expected to subtract neatly from one another.

The Māori Mathematics trajectory is not negative throughout the legacy series. Through much of the early and middle period, students identified as Māori are near the 2012 baseline and often slightly above the non-Māori comparison. That changes after 2018. The Māori line moves below baseline in 2019, falls further through the COVID-affected assessment points, and reaches its lowest legacy point in 2022.

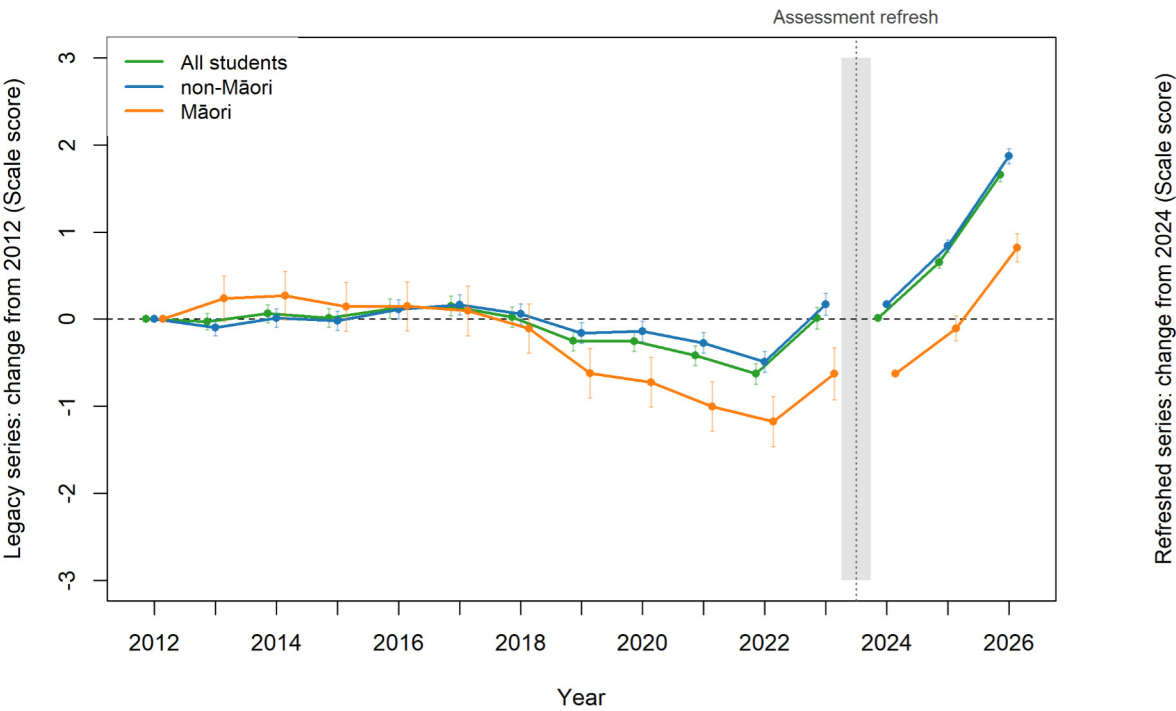


FIGURE 7a. **Māori Mathematics trajectory**

By 2023, students identified as Māori had partly recovered from the 2022 COVID-era trough, but the legacy endpoint remained below baseline at -0.63 [-0.93, -0.33]. The non-Māori comparison ended above baseline at 0.17 [0.04, 0.30]. The trajectory plot therefore shows a late-series shift: Māori students move from being near or slightly above the comparison trajectory in the mid-2010s to being below it by the end of the Mathematics legacy series.

The refreshed Mathematics segment shows positive movement for both groups. From the 2024 refreshed baseline to 2026, Māori students rose 1.45 [1.29, 1.61], while the non-Māori comparison rose 1.71 [1.62, 1.79]. This should be read as within-refreshed-series movement, not as a bridged estimate of change from 2023 to 2024.

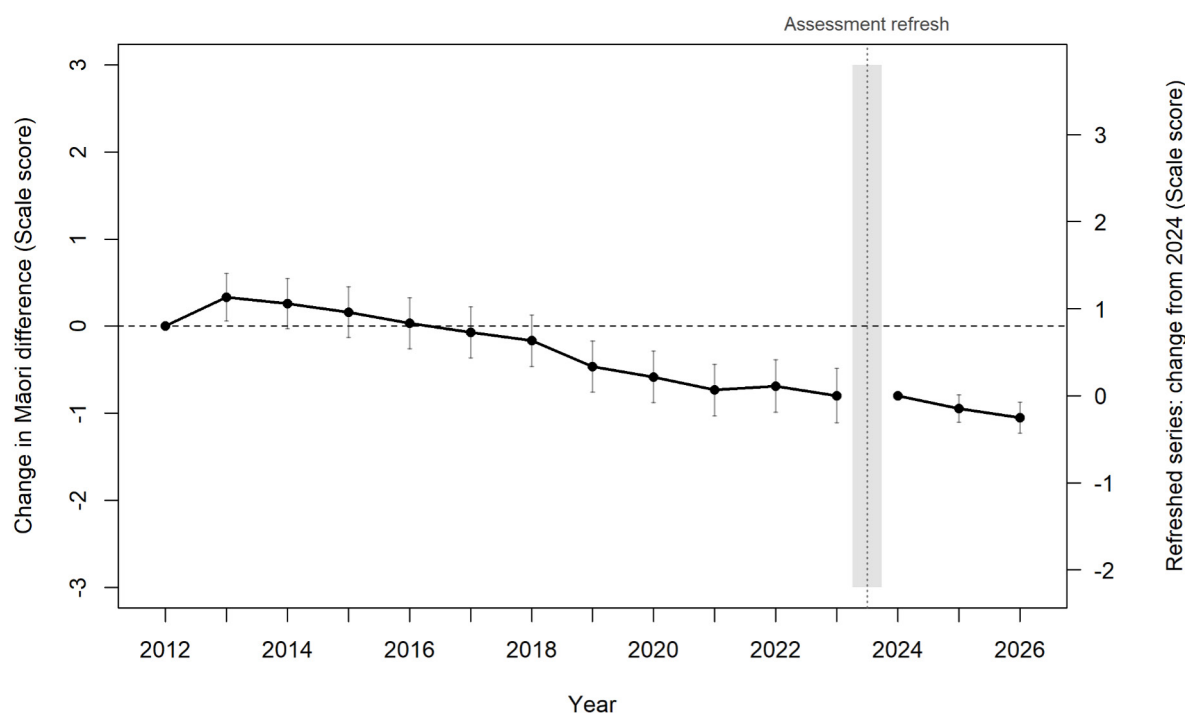


FIGURE 7b. **Māori relative to not-Māori same-starting-mix difference-over-time for PAT Mathematics**

The common-reference Mathematics figure strengthens the interpretation of uneven relative movement. In the legacy segment, the final Māori relative to non-Māori change estimate is -0.80 [-1.11, -0.49] in 2023. In the refreshed segment, the 2026 estimate is -0.25 [-0.43, -0.08] relative to the 2024 refreshed baseline.

The refreshed gap movement is smaller than the legacy movement, but it is in the same direction. This means Māori Mathematics improved strongly within the refreshed series, but not quite as much as the non-Māori comparison.

Māori Mathematics weakened after 2018 and recovered only partly by 2023. The refreshed trajectory was positive, but the Māori relative to non-Māori common-reference difference moved negatively in both the legacy and refreshed segments.

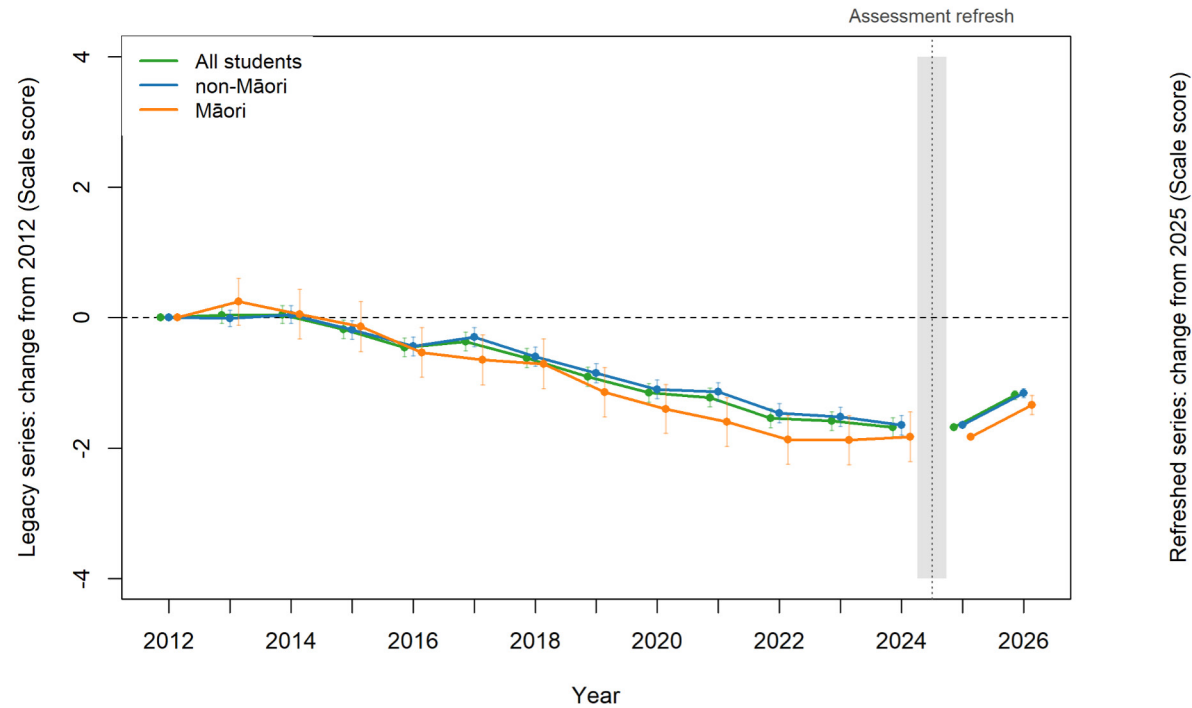


FIGURE 7c. Māori Reading Comprehension focal trajectory

For PAT Reading Comprehension, students identified as Māori ended the legacy segment below their own 2012 baseline at -1.82 [-2.21, -1.44]. The non-Māori comparison also declined, ending at -1.65 [-1.80, -1.50]. The focal estimate is slightly more negative, but the confidence intervals overlap and the broader pattern is a shared legacy decline.

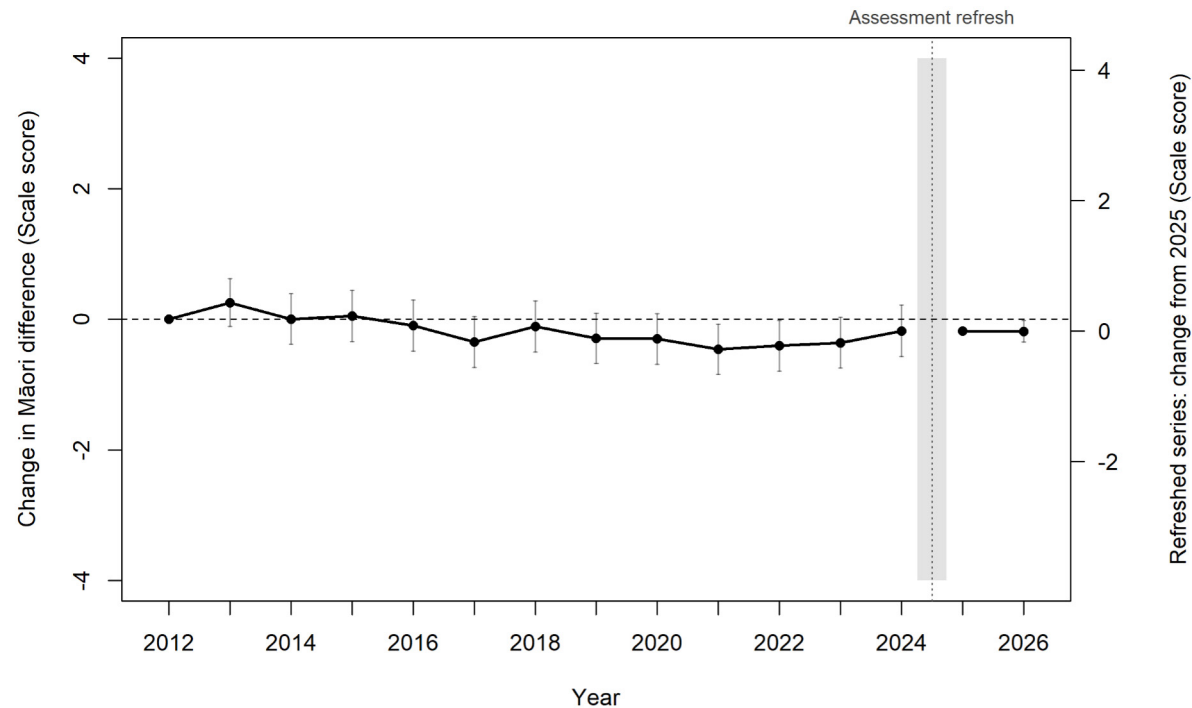


FIGURE 7d. Māori relative to non-Māori common-reference difference-over-time for PAT Reading Comprehension

The common-reference Reading Comprehension estimate is -0.18 [-0.57, 0.22] at the 2024 legacy endpoint, which is not statistically different from 0. In the refreshed segment, both Māori and non-Māori students move up by about the same amount from 2025 to 2026, with estimates of 0.49 [0.34, 0.64] for Māori and 0.49 [0.42, 0.56] for non-Māori. The refreshed common-reference gap estimate is effectively zero at 0.00 [-0.17, 0.16].

The practical implication is that PAT Reading Comprehension declined broadly in the legacy series, and the first refreshed comparison from 2025 to 2026 shows similar positive movement for Māori and non-Māori students. The evidence does not point to a distinctive Māori–non-Māori trajectory in Reading Comprehension once the common-reference comparison is applied. The main issue is the broad reading comprehension decline discussed throughout the report.

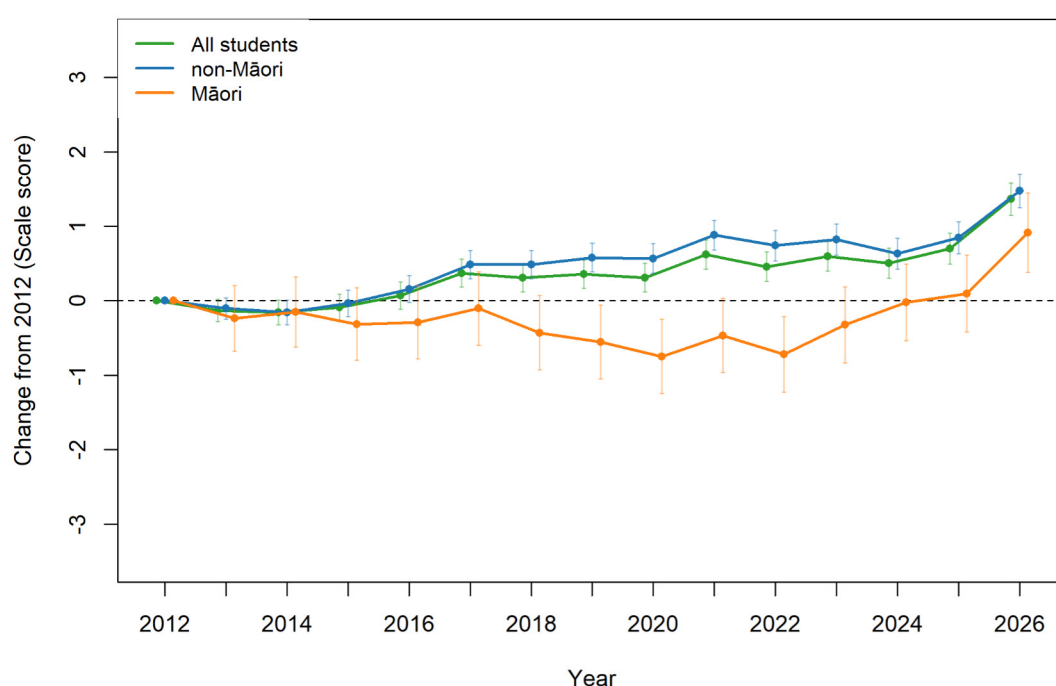


FIGURE 7e. **Māori Reading Vocabulary focal trajectory**

For PAT Reading Vocabulary, students identified as Māori ended above baseline in 2026, at 0.91 [0.38, 1.45]. This contrasts with the PAT Reading Comprehension legacy decline.

The shape of the line matters. The Māori Vocabulary trajectory is near baseline in the early years, then falls below baseline across parts of the late 2010s and early 2020s. It turns slightly positive in 2024 and 2025, then rises more clearly in 2026. The non-Māori comparison follows a steadier positive path and ends higher, at 1.47 [1.25, 1.70].

Māori Reading Vocabulary ended above baseline, but the increase appeared later and was smaller than for the non-Māori comparison.

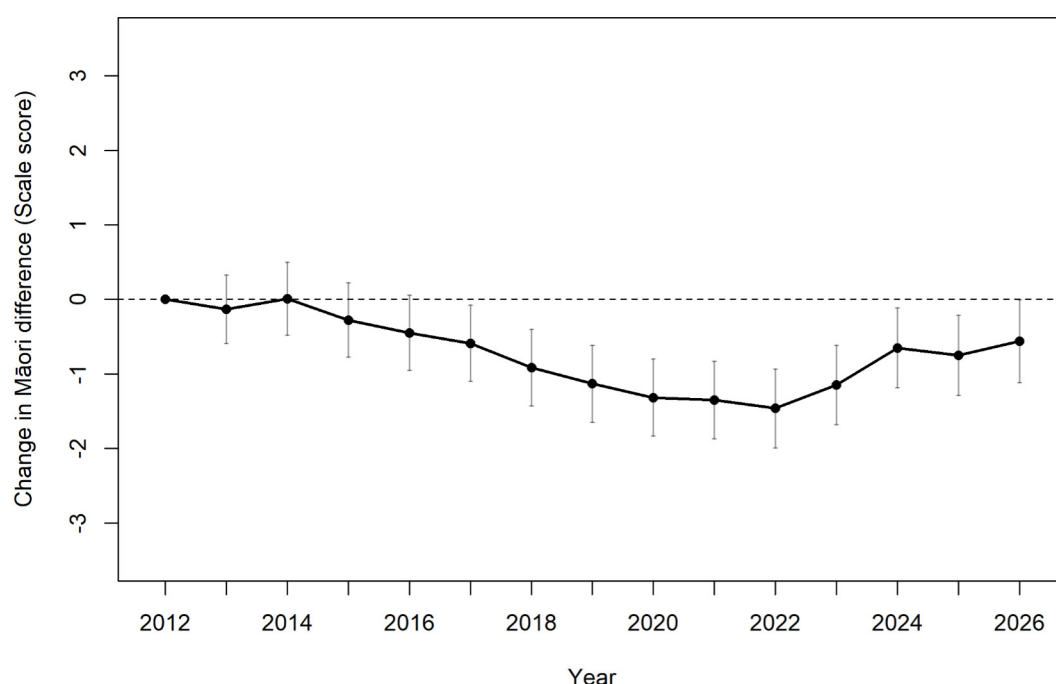


FIGURE 7f. **Māori relative to non-Māori common-reference difference-over-time for PAT Reading Vocabulary**

The common-reference Reading Vocabulary output qualifies the positive focal trajectory. The final Māori relative to non-Māori common-reference change estimate is -0.56 [-1.12, -0.01]. This does not mean Māori Reading Vocabulary declined. The trajectory plot shows the opposite: Māori Reading Vocabulary ends above baseline. It means that, under the common-reference comparison, the Māori–non-Māori difference remained negative in point-estimate terms for much of the series, and the 2026 interval now narrowly excludes zero.

Taken together, the Māori profile is mixed across assessment areas and segments. Mathematics shows the clearest less favourable relative movement, as the legacy endpoint is below baseline for Māori students, and the common-reference gap is negative in both the legacy and refreshed segments, even though the refreshed Mathematics trajectory itself is strongly positive. Reading Comprehension is mainly a broad decline story, with no clear Māori–non-Māori gap movement in either the legacy or refreshed segment. Reading Vocabulary ends positively for Māori students, but the uplift is later and less strong than for non-Māori students.

9.4 Pacific²

This subsection explores how PAT assessment scores for students identified as Pacific have changed over time.

2 Analytically, the Pacific category used in this report was constructed using total-response ethnicity indicators for Samoan, Cook Islands Māori, Tongan, Niuean, Tokelauan, Fijian, and Other Pacific Peoples. Students could also be represented in another ethnicity category. Government terminology for this grouping has varied across agencies and over time. Terms have included Pacific, Pacific peoples, Pacific Island/Pacific Islanders, Pasifika, and non-English forms such as Pasefika and Pasifiki. The Ministry for Pacific Peoples generally uses Pacific people or Pacific peoples, while earlier education policy commonly used Pasifika—for example, the Pasifika Education Plan was replaced by the Action Plan for Pacific Education in 2020. For consistency, this report uses Pacific as the analytic label and Pacific students in prose. This umbrella category encompasses diverse ethnic, national, linguistic, and multi-ethnic identities and should not be understood as a homogeneous population.

As will be discussed in more detail below, the Pacific Mathematics legacy trajectory ends close to baseline after the COVID period dip, and the refreshed Mathematics segment rises strongly from 2024 to 2026. Pacific Reading Comprehension declines in the legacy series, but less than the non-Pacific comparison. The refreshed Reading Comprehension segment is positive for both groups, though slightly less positive for Pacific students. Reading Vocabulary is positive by 2026, but the common-reference output remains cautious about reading this as a distinct Pacific-specific relative shift.

TABLE 10. Pacific companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final comparison trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	-0.14 [-0.54, 0.25]	0.03 [-0.09, 0.15]	-0.18 [-0.58, 0.23]
Mathematics refreshed	2024 to 2026	1.46 [1.24, 1.67]	1.68 [1.59, 1.76]	-0.22 [-0.45, 0.01]
Reading Comprehension legacy	2012 to 2024	-1.06 [-1.55, -0.57]	-1.76 [-1.91, -1.61]	0.70 [0.21, 1.19]
Reading Comprehension refreshed	2025 to 2026	0.30 [0.10, 0.49]	0.52 [0.45, 0.58]	-0.22 [-0.42, -0.01]
Reading Vocabulary	2012 to 2026	1.54 [0.87, 2.21]	1.34 [1.12, 1.56]	0.20 [-0.48, 0.88]

The Pacific Mathematics trajectory remains close to baseline through much of the legacy series, dips in 2021 and especially 2022, and then returns close to baseline in 2023. At the 2023 legacy endpoint, students identified as Pacific are at -0.14 [-0.54, 0.25], while the non-Pacific comparison is at 0.03 [-0.09, 0.15]. The two endpoints are close, and the confidence intervals overlap. The result fits the broader Mathematics pattern of a lowest point during the COVID-19 period followed by initial recovery.

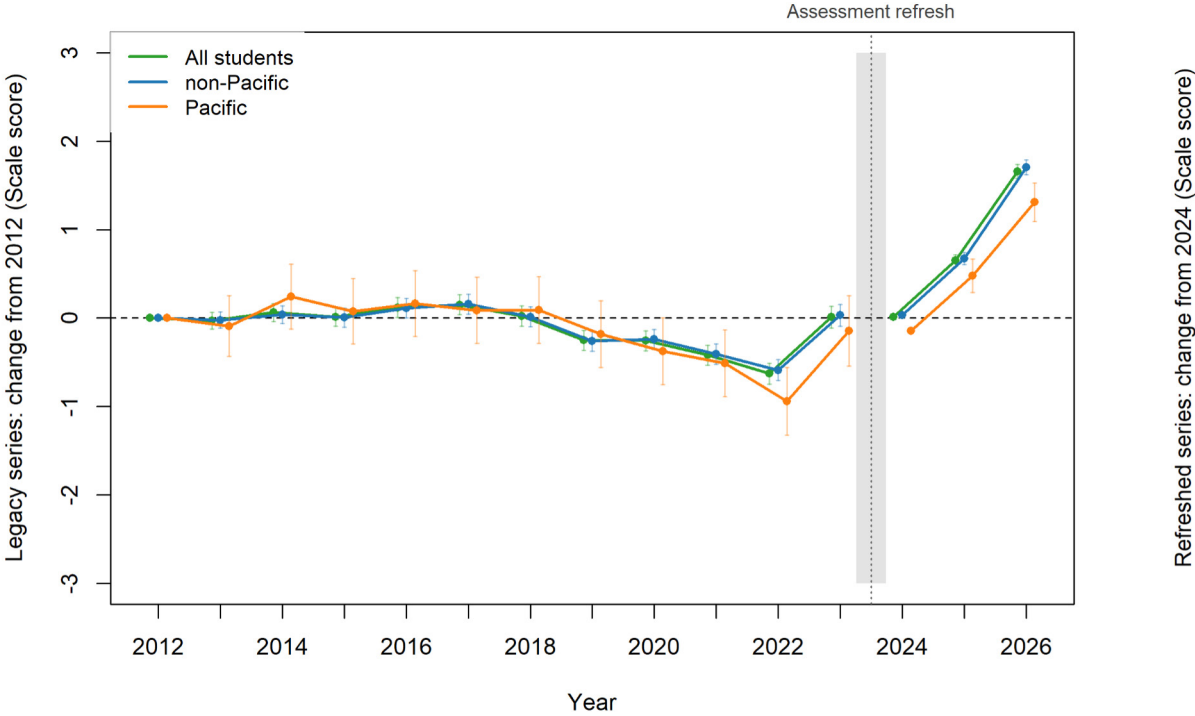


FIGURE 8a. Pacific Mathematics focal trajectory

The refreshed Mathematics segment is positive for both groups. From the 2024 refreshed baseline to 2026, Pacific students increased 1.46 [1.24, 1.67], while the non-Pacific comparison increased 1.68 [1.59, 1.76]. This should be read as within refreshed series movement, not as a bridged estimate of change from 2023 to 2024.

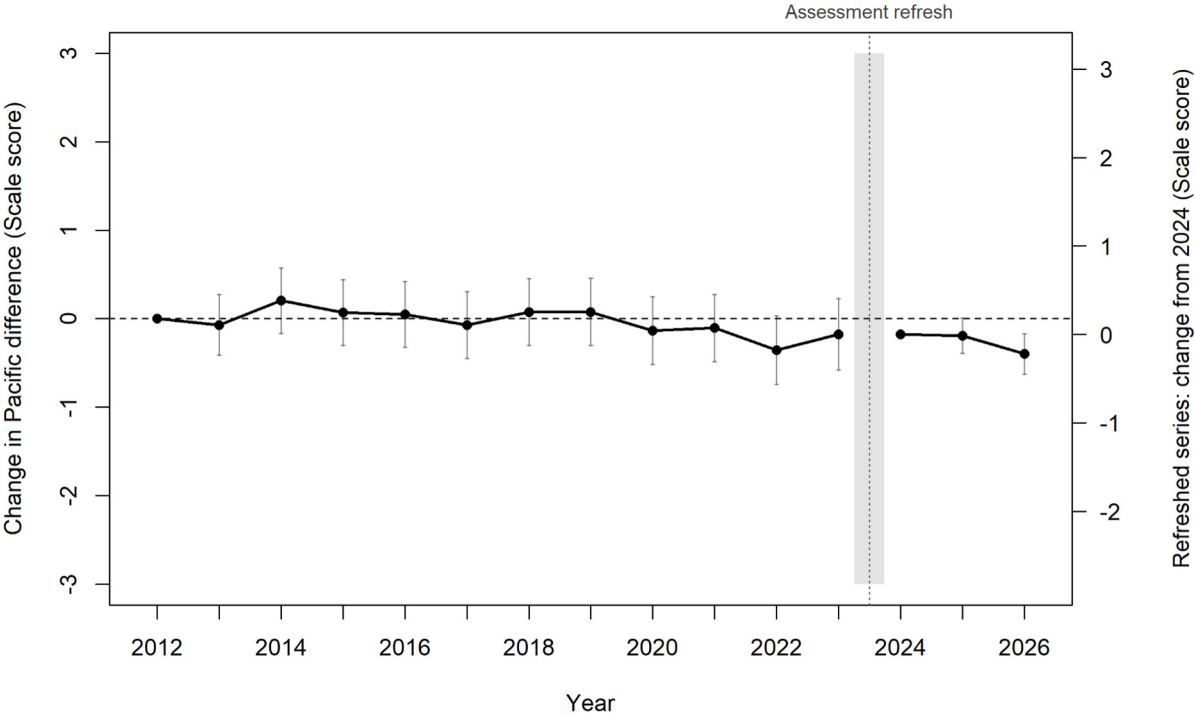


FIGURE 8b. Pacific relative to non-Pacific common-reference difference-over-time for PAT Mathematics

The common-reference Mathematics figure suggests that the relative Pacific movement is not more favourable than the comparison movement. In the legacy segment, the final Pacific relative to non-Pacific change estimate is -0.18 $[-0.58, 0.23]$, which gives little evidence of a Pacific-specific Mathematics shift across the legacy series. In the refreshed segment, the 2026 estimate is -0.22 $[-0.45, 0.01]$ relative to the 2024 refreshed baseline. Both groups improve and the comparison-group point estimate rises slightly more, but the refreshed relative interval includes zero.

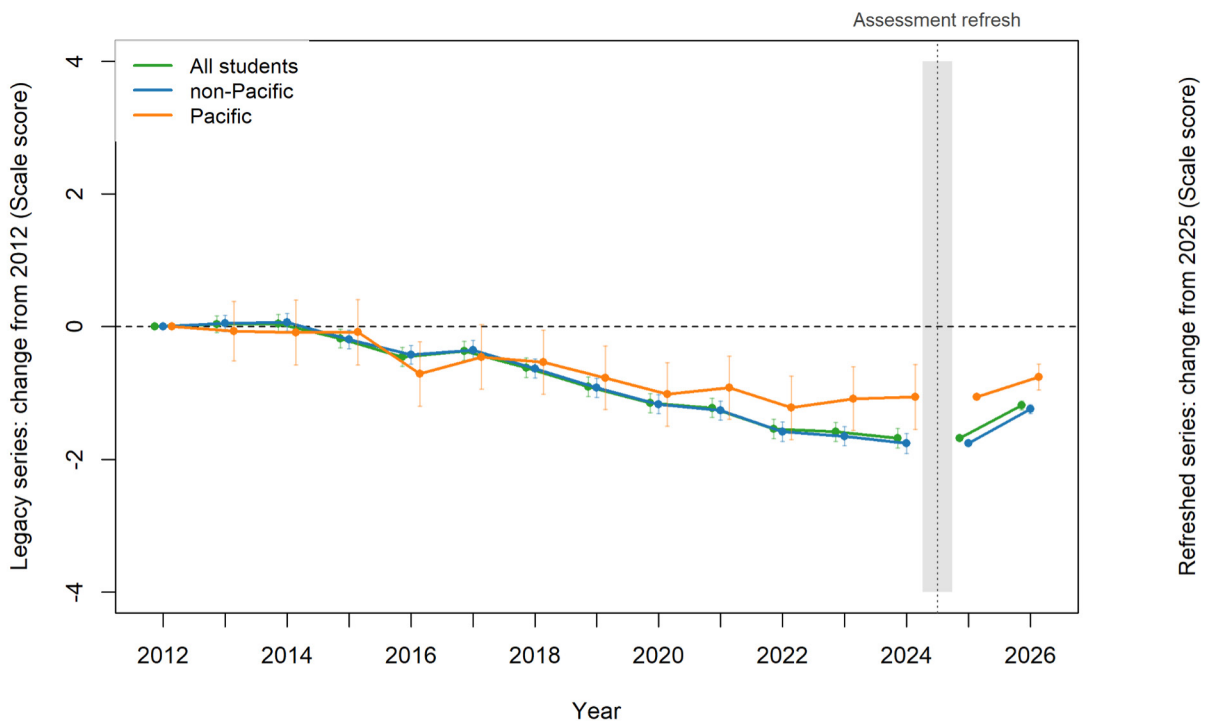


FIGURE 8c. Pacific Reading Comprehension focal trajectory

The story for Pacific Reading Comprehension retains the earlier central finding. Students identified as Pacific decline in the legacy series, ending at -1.06 $[-1.55, -0.57]$, so this is not a positive achievement outcome in absolute terms. However, the non-Pacific comparison declines more, ending at -1.76 $[-1.91, -1.61]$. The common reference gap change at the 2024 legacy endpoint is 0.70 $[0.21, 1.19]$, supporting the view that Pacific Reading Comprehension declined less than the comparison group through the legacy series.

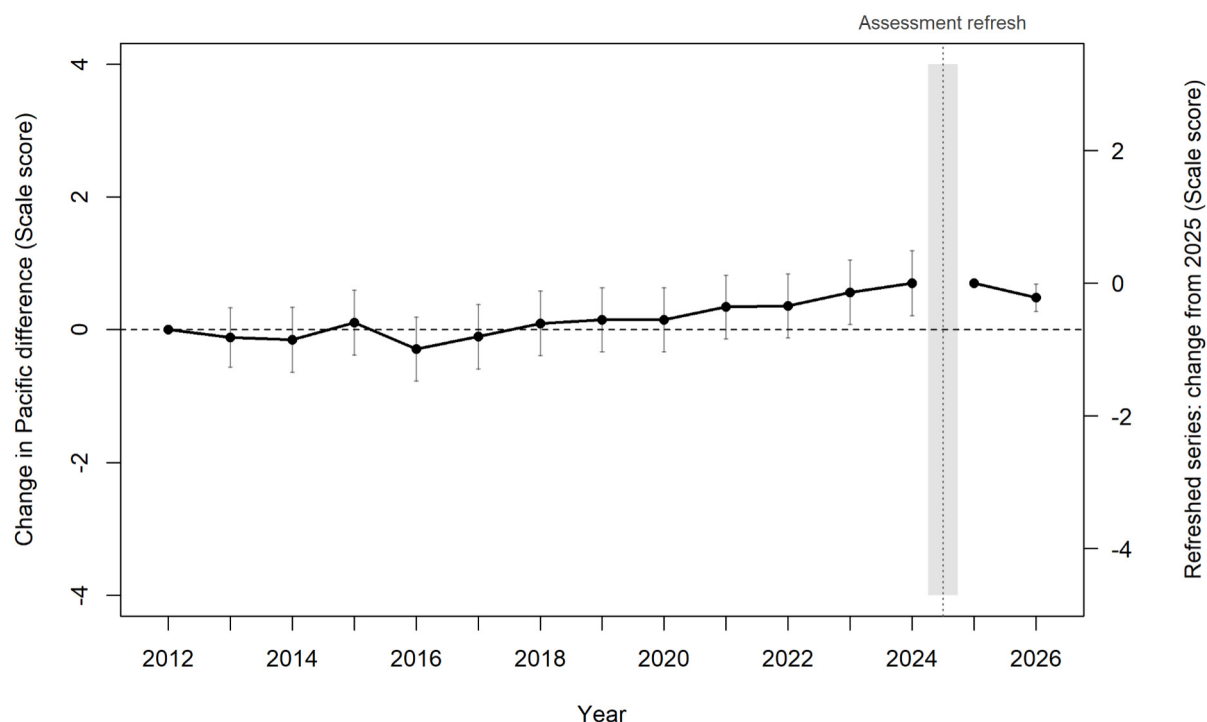


FIGURE 8d. **Pacific relative to not-Pacific same-starting-mix difference-over-time for PAT Reading Comprehension**

From 2025 to 2026, the refreshed Reading Comprehension segment rises for both groups. Pacific students increase 0.30 [0.10, 0.49], while the non-Pacific comparison increases 0.52 [0.45, 0.58]. The common reference gap change is -0.22 [-0.42, -0.01], so the relative movement in the first refreshed comparison is slightly less favourable for Pacific students. This does not undo the legacy finding, though it does show that the positive relative legacy pattern should not be projected automatically into the refreshed series.

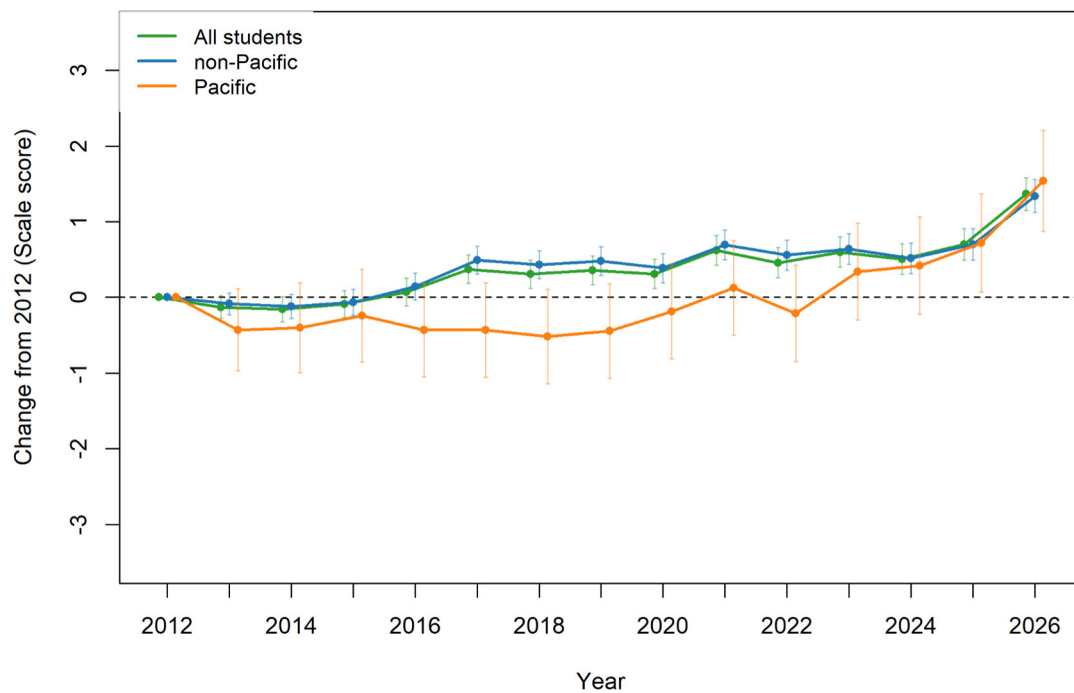


FIGURE 8e. **Pacific Reading Vocabulary focal trajectory**

The Pacific PAT Reading Vocabulary trajectory is positive by 2026. Students identified as Pacific end above baseline at 1.54 [0.87, 2.21]. The non-Pacific comparison also ends above baseline, at 1.34 [1.12, 1.56]. Both estimates are clearly positive. The Pacific point estimate is higher, but the confidence intervals overlap and the line is variable across earlier years.

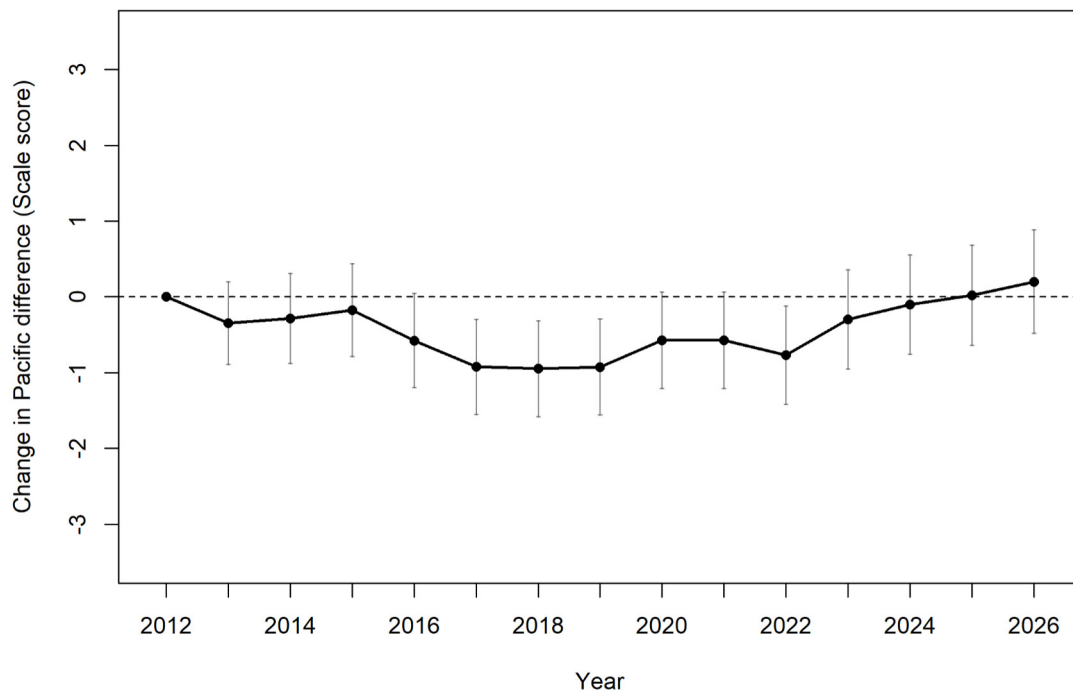


FIGURE 8f. **Pacific relative to non-Pacific common-reference difference-over-time for PAT Reading Vocabulary**

The final Pacific relative to non-Pacific common-reference change estimate is 0.20 [-0.48, 0.88]. This means there is not strong evidence that the 2026 Pacific Vocabulary endpoint reflects a distinct Pacific-specific widening in relative terms. The safest reading is that Pacific students participated in the wider Reading Vocabulary uplift, rather than that Reading Vocabulary changed uniquely for Pacific students.

Overall, the Pacific profile is mixed. In Mathematics, the legacy series recovers close to baseline, and the refreshed segment improves strongly. In Reading Comprehension, the legacy decline is less for Pacific students than for the comparison group, which remains the clearest relative finding in this subsection. In Reading Vocabulary, Pacific students show strong positive movement by 2026, but the common-reference gap is not clearly different from zero. The central message is relative resilience in legacy Reading Comprehension and strong participation in the Reading Vocabulary uplift, rather than a simple story of improvement or decline.

9.5 European/Pākehā

For European/Pākehā students, the consistent feature across the three PAT assessment series considered in this report is negative relative movement. In the legacy Mathematics and Reading Comprehension segments, students identified as European/Pākehā end below their own 2012 baseline and move less favourably than the non-European/Pākehā comparison. In the refreshed Mathematics and Reading Comprehension segments, both the focal and non-European/Pākehā trajectories rise, but the common-reference gap change is again less favourable for European/Pākehā in Mathematics and close to zero in Reading Comprehension. In Reading Vocabulary, European/Pākehā students end above baseline, but less than the non-European/Pākehā comparison.

For Mathematics and Reading Comprehension, the focal trajectory and common-reference figures should be read within assessment series. They show movement within the legacy segment and movement within the refreshed segment, not a numerical estimate of cross-boundary change.

TABLE 11. European/Pākehā companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final non-European/ Pākehā trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	-0.39 [-0.58, -0.19]	0.56 [0.35, 0.77]	-0.94 [-1.27, -0.62]
Mathematics refreshed	2024 to 2026	1.54 [1.44, 1.65]	1.80 [1.66, 1.93]	-0.25 [-0.44, -0.07]
Reading Comprehension legacy	2012 to 2024	-2.09 [-2.34, -1.85]	-1.11 [-1.36, -0.86]	-0.98 [-1.39, -0.58]
Reading Comprehension refreshed	2025 to 2026	0.47 [0.37, 0.56]	0.53 [0.41, 0.64]	-0.06 [-0.23, 0.11]
Reading Vocabulary	2012 to 2026	1.02 [0.68, 1.36]	1.88 [1.51, 2.26]	-0.86 [-1.43, -0.29]

The focal and non-European/Pākehā estimates come from the trajectory plots. The common-reference gap change estimate shows change in the European/Pākehā relative to non-European/Pākehā difference from the relevant segment baseline when both groups are evaluated against the same reference population. For Reading Vocabulary, the single segment baseline is 2012.

The Mathematics trajectory shows a late series weakening for students identified as European/Pākehā. The focal line is close to baseline through 2017, then moves below baseline from 2018 onward. It reaches its lowest point during the COVID-19 period in 2022, reflecting learning from 2021, before a partial recovery in 2023.

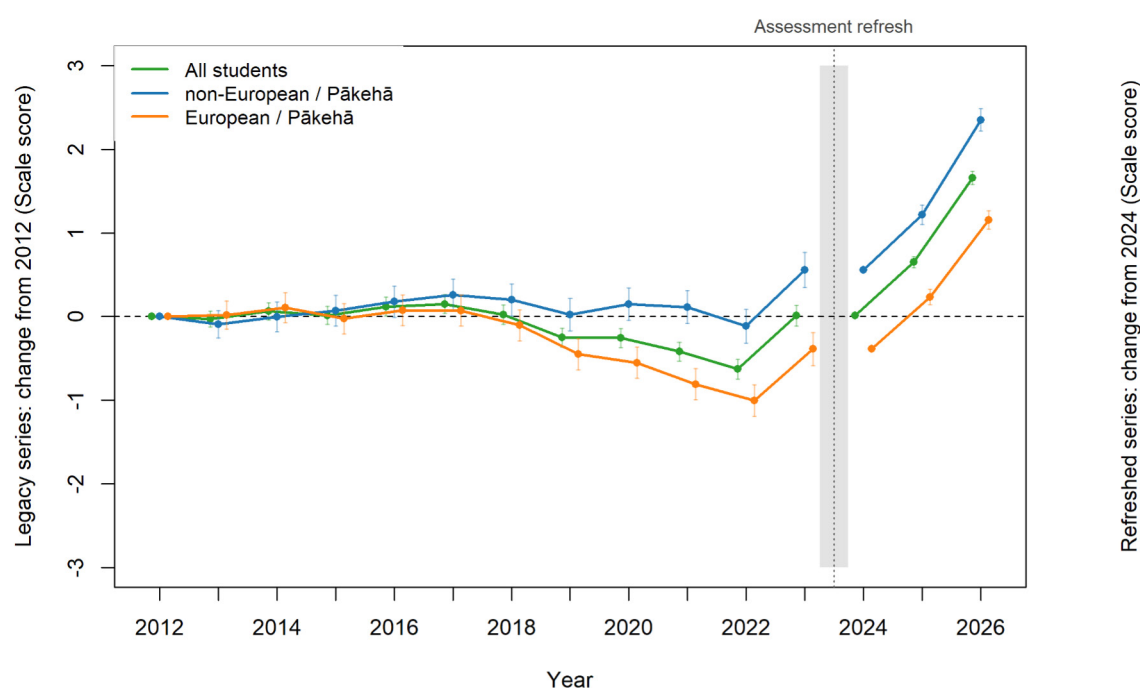


FIGURE 9a. European/Pākehā Mathematics focal trajectory

By the final legacy Mathematics year, students identified as European/Pākehā are below their 2012 baseline at -0.39 [-0.58, -0.19]. The non-European/Pākehā comparison is above baseline at 0.56 [0.35, 0.77]. It is important to keep in mind that this should not be read as an absolute ranking of ethnic groups. The plot shows change from each group's own 2012 starting point, and not all ethnic groups started in the same place in 2012.

Within the refreshed Mathematics segment, European/Pākehā students rise 1.54 [1.44, 1.65] from the 2024 refreshed baseline by 2026. The non-European/Pākehā comparison also rises, at 1.80 [1.66, 1.93]. The refreshed pattern is therefore positive, but the relative change remains less favourable for European/Pākehā.

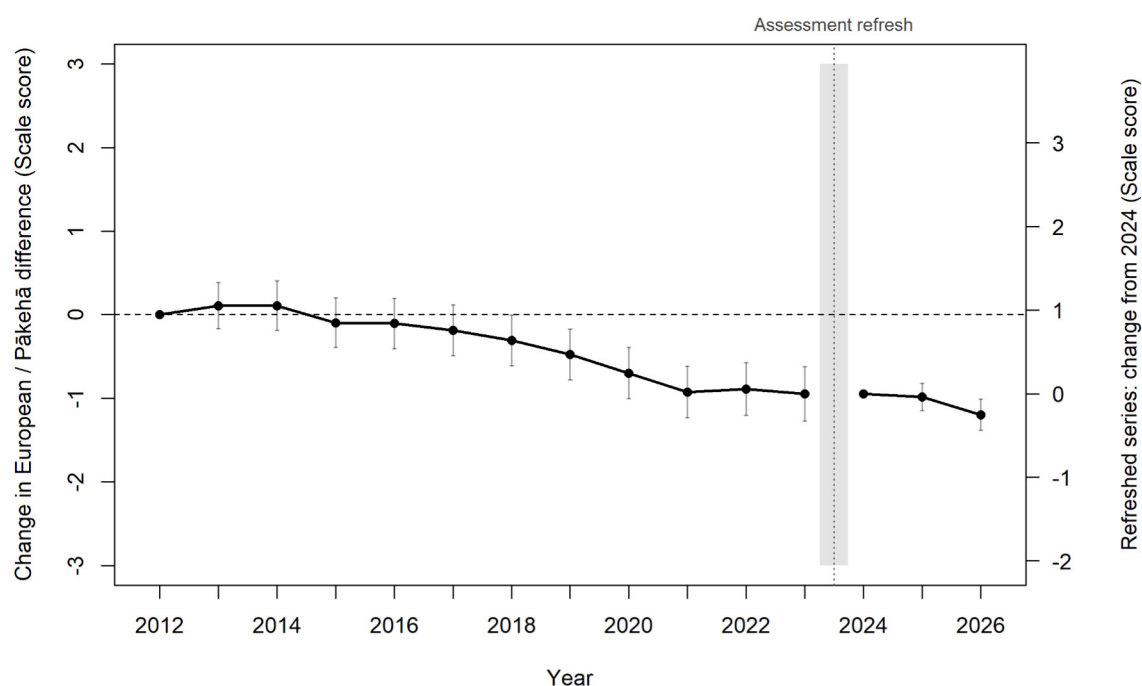


FIGURE 9b. **European/Pākehā relative to not European/Pākehā common-reference difference-over-time for PAT Mathematics**

The common-reference Mathematics output supports the negative relative interpretation. The final legacy European/Pākehā relative to non-European/Pākehā change estimate is -0.94 [-1.27, -0.62]. The refreshed estimate is -0.25 [-0.44, -0.07] by 2026. In other words, this PAT Mathematics shift is not just a feature of the wider makeup of the European/Pākehā and non-European/Pākehā groups. The focal contrast itself moved negatively in both segments.

The education interpretation is that Mathematics weakening was not limited to groups usually foregrounded in equity discussions. European/Pākehā students were a major part of the legacy Mathematics downturn, and their refreshed improvement did not remove the negative relative pattern.

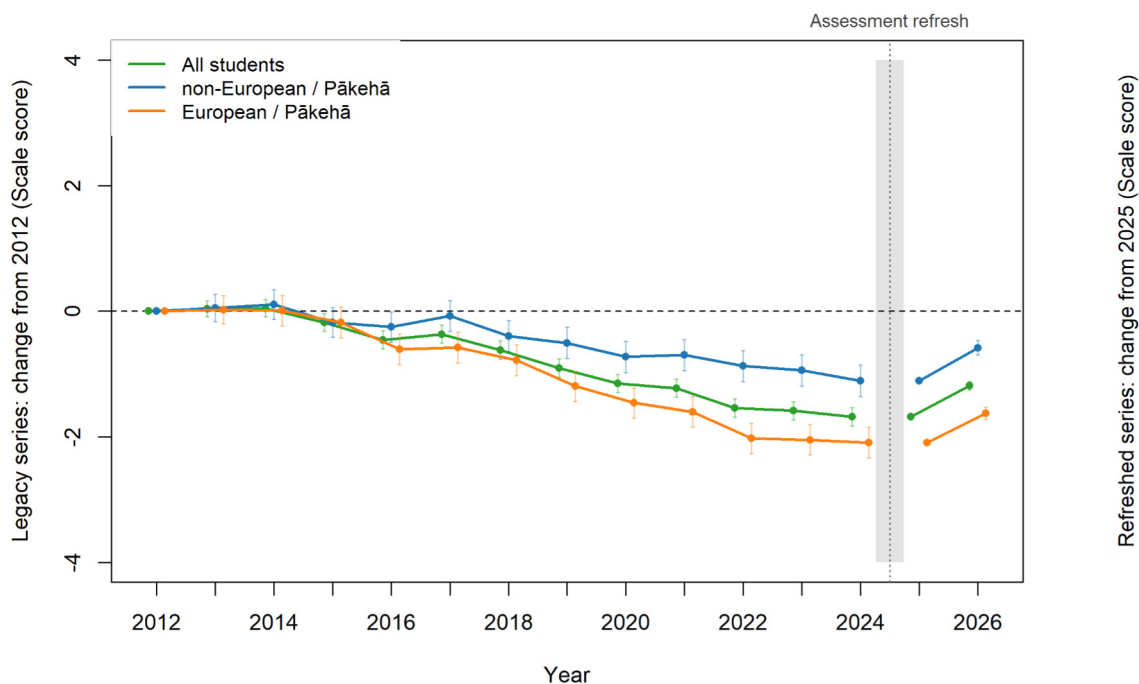


FIGURE 9c. **European/Pākehā Reading Comprehension focal trajectory**

Reading Comprehension is the strongest negative focal result in this profile. Students identified as European/Pākehā end the legacy segment at -2.09 [-2.34, -1.85], while the non-European/Pākehā comparison ends at -1.10 [-1.36, -0.85]. Both trajectories are below baseline, but the European/Pākehā trajectory falls further from its own 2012 starting point.

This is not a short-term COVID-19 dip. The European/Pākehā legacy trajectory is already below baseline by the mid-2010s and declines from 2017 until the end of the legacy series. This fits the broader Reading Comprehension pattern, where decline begins before the pandemic years and continues to the end of the legacy series.

Within the refreshed Reading Comprehension segment, the focal and non-European/Pākehā lines both rise from 2025 to 2026. The European/Pākehā estimate is 0.47 [0.37, 0.56], while the non-European/Pākehā comparison is 0.53 [0.41, 0.64].

European/Pākehā students show one of the clearest downward legacy Reading Comprehension trajectories among the focal groups. The decline is therefore not confined to groups commonly described as underserved.

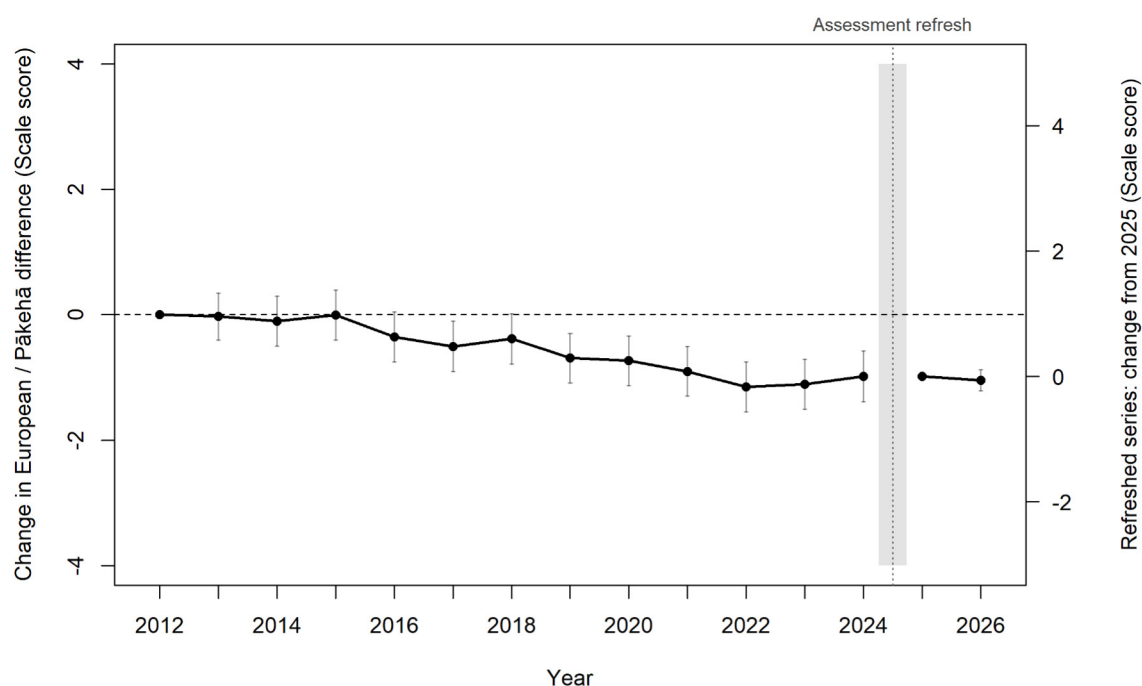


FIGURE 9d. **European/Pākehā relative to not European/Pākehā same-starting-mix difference-over-time for PAT Reading Comprehension**

The common-reference Reading Comprehension output strengthens the same interpretation. The final legacy change estimate is -0.98 $[-1.39, -0.58]$. This means that, when European/Pākehā and non-European/Pākehā students are compared using the same reference population, the European/Pākehā relative to non-European/Pākehā difference still moves clearly in a less favourable direction through the legacy segment. In the refreshed segment, the 2026 common-reference estimate is -0.06 $[-0.23, 0.11]$, so the shorter refreshed series does not show a clear additional relative movement.

This is stronger than saying that the focal trajectory declined. The common-reference comparison shows clear negative relative movement through the legacy series. The short-refreshed series does not show a clear further change in that relative position.

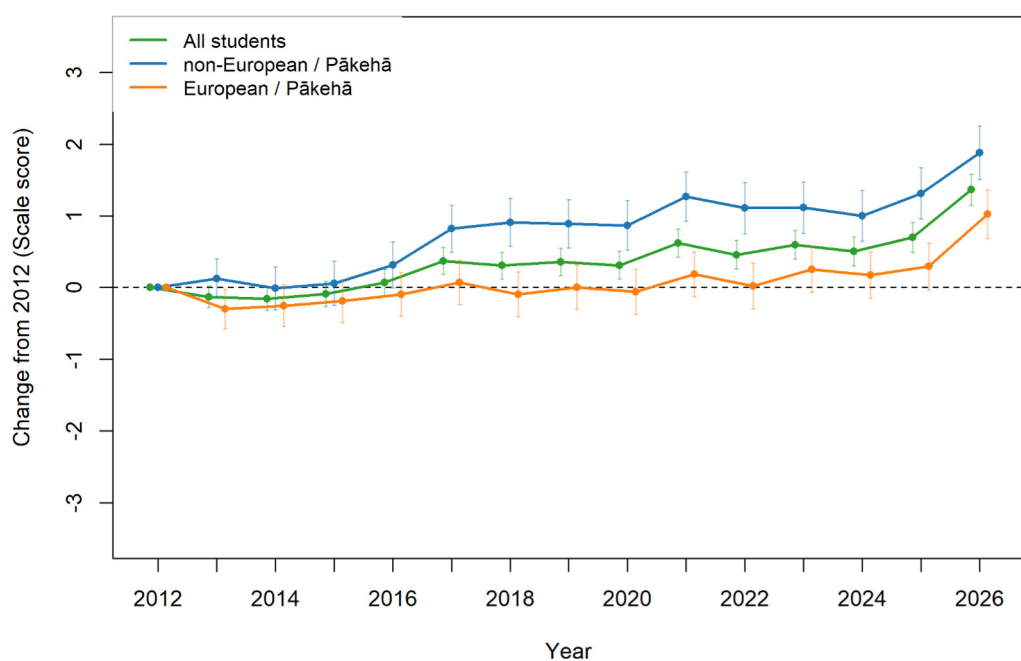


FIGURE 9e. **European/Pākehā Reading Vocabulary focal trajectory**

Students identified as European/Pākehā end above baseline in Reading Vocabulary at 1.02 [0.68, 1.36]. Students who do not identify as European/Pākehā also end above baseline, and more strongly, at 1.88 [1.51, 2.26].

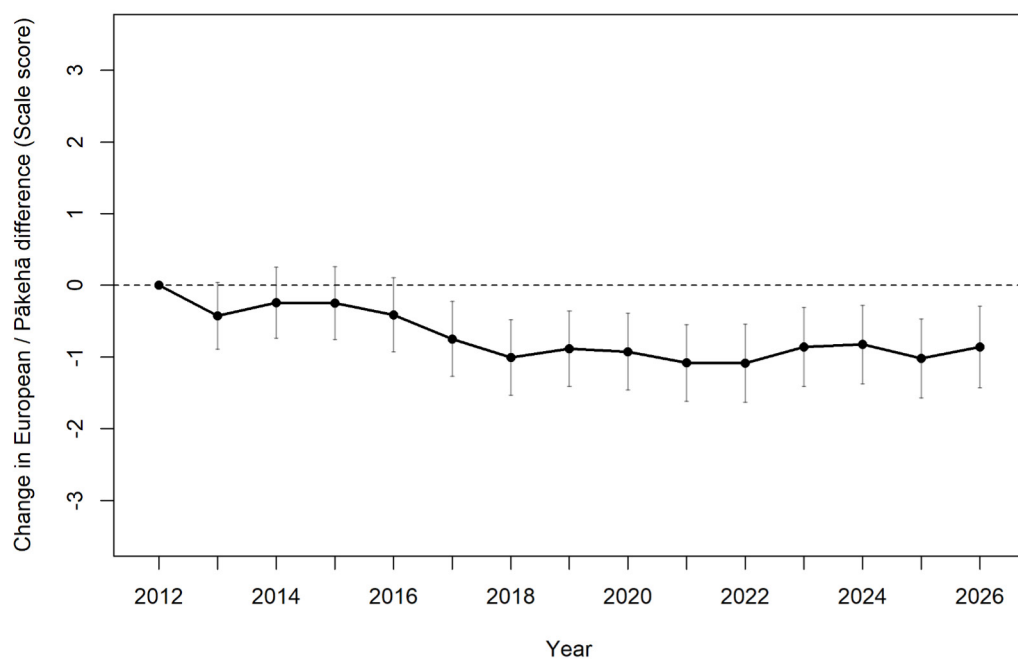


FIGURE 9f. **European/Pākehā relative to not European/Pākehā same-starting-mix difference-over-time for PAT Reading Vocabulary**

The common-reference Reading Vocabulary figure confirms the basic story in the trajectory plot. The final European/Pākehā relative to non-European/Pākehā change estimate is -0.86 [-1.43, -0.29]. This does not mean European/Pākehā Reading Vocabulary declined. It means that, under a common-reference comparison, European/Pākehā students gained less in vocabulary than those who did not identify as European/Pākehā.

European/Pākehā results show negative relative movement in all three measures. Legacy Reading Comprehension is the clearest case; Mathematics rises in the refreshed series but remains less favourable in relative terms, and Vocabulary rises less than in the comparison group. This shows why European/Pākehā students should be analysed rather than treated only as a default comparison group.

9.6 North Asian

North Asian results show a strong upward PAT Mathematics trajectory, a positive PAT Reading Vocabulary trajectory, and a Reading Comprehension trajectory that declined in the legacy segment and moved less in the refreshed segment than the non-North Asian comparison. The common-reference plots are especially important here because they sharpen the interpretation of PAT Mathematics and PAT Reading Comprehension.

For Mathematics and Reading Comprehension, the focal trajectory and common-reference figures should be read within assessment series. They show movement within the legacy segment and movement within the refreshed segment, not a numerical estimate of change across the assessment refresh.

TABLE 12. North Asian companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final non-North Asian trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	0.89 [0.42, 1.36]	-0.08 [-0.20, 0.05]	0.97 [0.49, 1.44]
Mathematics refreshed	2024 to 2026	1.92 [1.67, 2.17]	1.62 [1.54, 1.71]	0.30 [0.03, 0.56]
Reading Comprehension legacy	2012 to 2024	-2.00 [-2.59, -1.42]	-1.64 [-1.79, -1.50]	-0.36 [-0.96, 0.23]
Reading Comprehension refreshed	2025 to 2026	0.21 [0.00, 0.42]	0.53 [0.46, 0.59]	-0.32 [-0.54, -0.09]
Reading Vocabulary	2012 to 2026	1.65 [0.81, 2.48]	1.34 [1.13, 1.56]	0.30 [-0.54, 1.14]

The North Asian Mathematics trajectory is one of the clearer positive results in the ethnicity chapter. In the final legacy Mathematics year, students identified as North Asian end above their 2012 baseline at 0.89 [0.42, 1.36]. In contrast, the non-North Asian comparison is close to baseline at -0.08 [-0.20, 0.05].

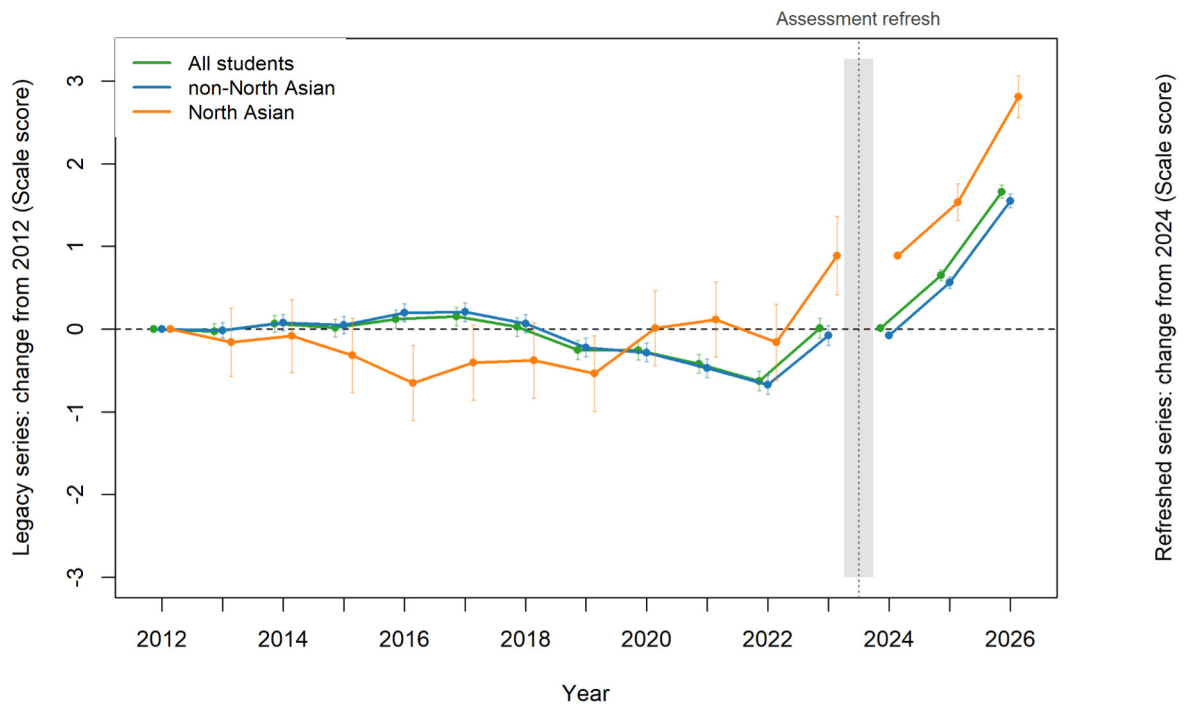


FIGURE 10a. North Asian Mathematics focal trajectory

It is important to note that the North Asian Mathematics trajectory line is not high throughout the legacy series. It is weaker in the mid-2010s and then rises in the early 2020s, reaching its highest legacy point in 2023. Within the refreshed Mathematics segment, both groups improve from 2024 to 2026. North Asian students rise 1.92 [1.67, 2.17], while the non-North Asian comparison rises 1.62 [1.54, 1.71].

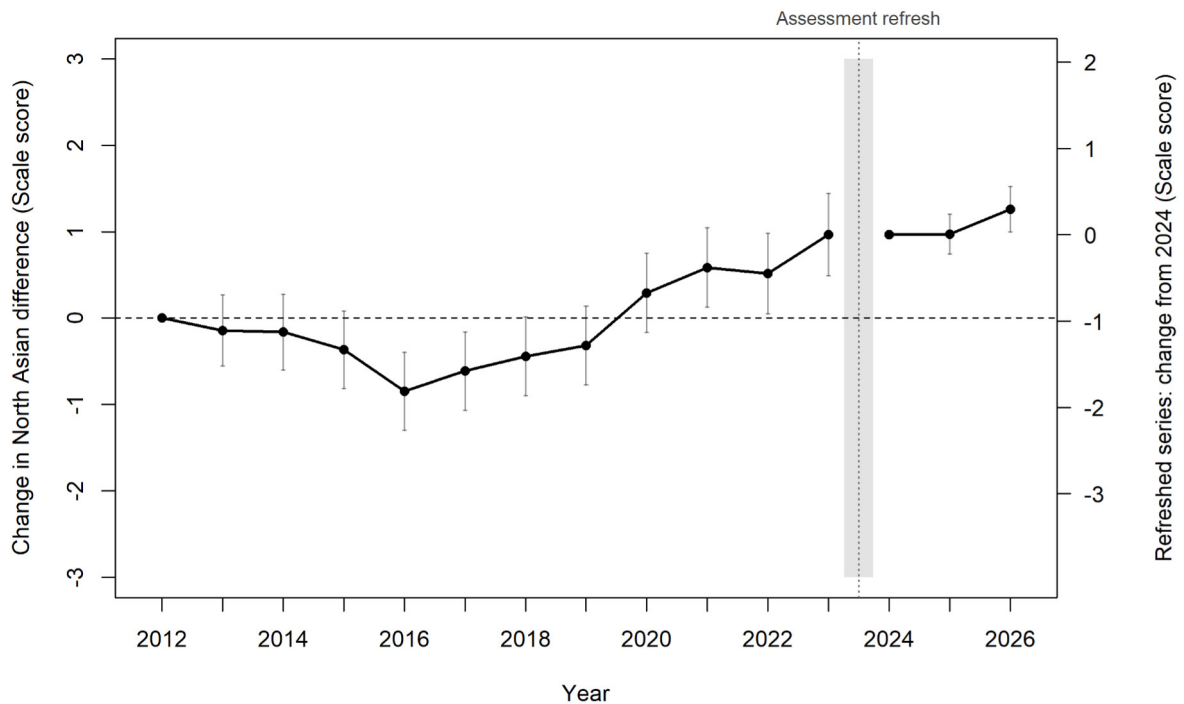


FIGURE 10b. North Asian relative to non-North Asian same common-reference difference-over-time for PAT Mathematics

The common-reference Mathematics output supports the positive relative interpretation in the legacy segment. The final North Asian relative to non-North Asian common-reference gap change estimate is 0.97 [0.49, 1.44] in 2023. This means that, under the common-reference comparison, the North Asian relative position in Mathematics moved about one scale-score point in a favourable direction from its 2012 value.

The refreshed gap change estimate is 0.30 [0.03, 0.56] by 2026. This indicates a modest widening in the North Asian direction during the refreshed Mathematics segment, while both the North Asian and comparison trajectories rose strongly from the refreshed baseline.

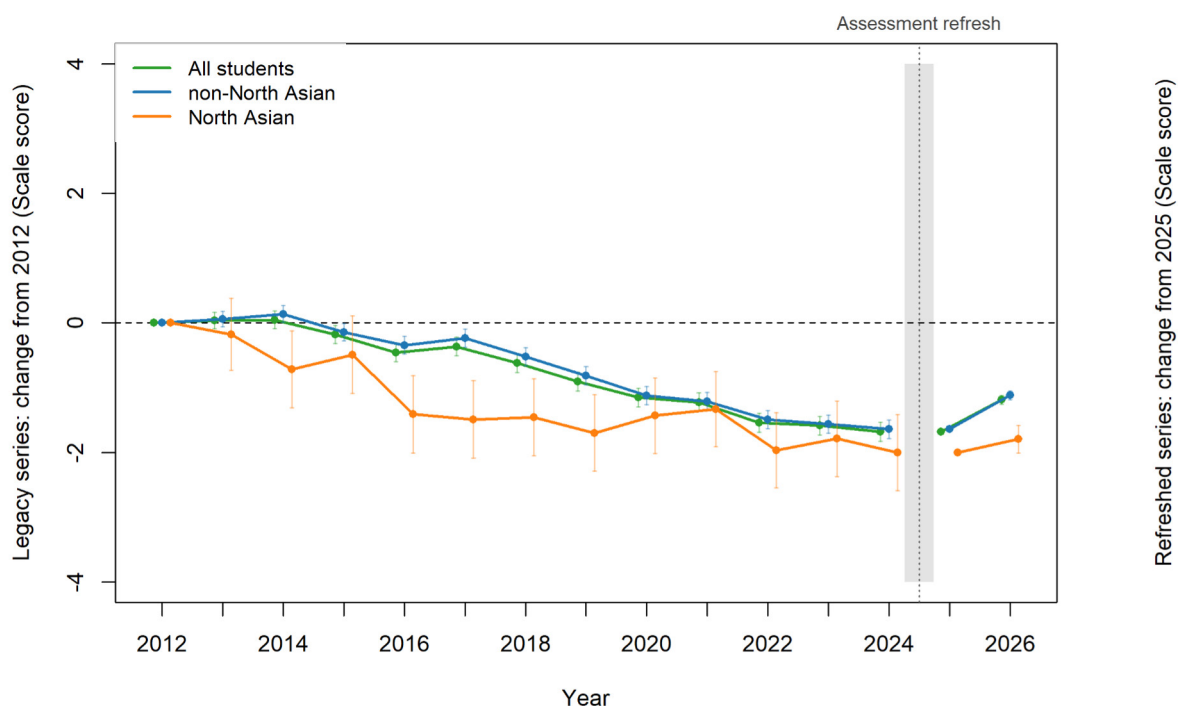


FIGURE 10c. North Asian Reading Comprehension focal trajectory

Reading Comprehension is much less positive. Students identified as North Asian end the legacy segment below baseline at -2.00 [-2.59, -1.42]. The non-North Asian comparison also ends below baseline, at -1.64 [-1.79, -1.50]. In the trajectory plot, the North Asian legacy line therefore finishes slightly more negatively than the comparison, but both lines are clearly below their own 2012 baselines.

North Asian Reading Comprehension also declined in the legacy series, consistent with the broad direction seen across most focal groups.

Within the refreshed Reading Comprehension segment, North Asian students rise 0.21 [-0.00, 0.42] from 2025 to 2026, while the non-North Asian comparison rises 0.53 [0.46, 0.59]. This early refreshed movement is positive in point estimate terms for North Asian students, but the interval crosses zero and the increase is smaller than for the non-North Asian comparison.

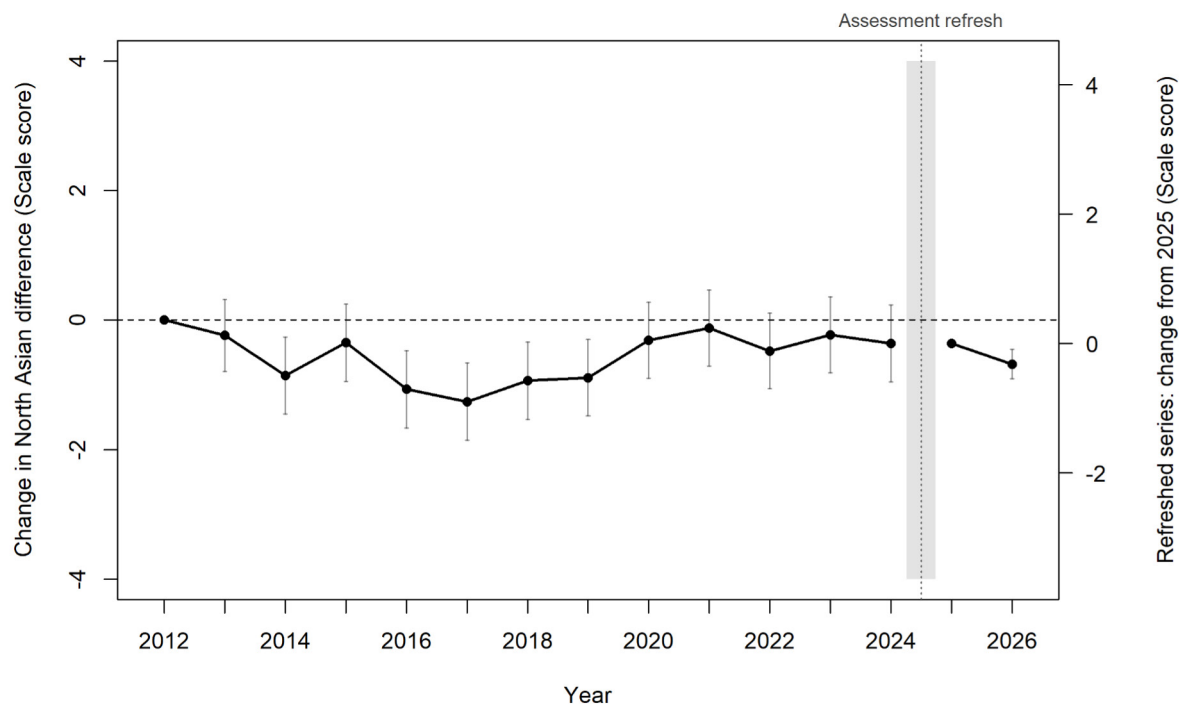


FIGURE 10d. North Asian relative to non-North Asian common-reference over time for PAT Reading Comprehension

The common-reference Reading Comprehension output makes the relative interpretation more cautious for the legacy segment and less favourable for the refreshed segment. The final legacy North Asian relative to non-North Asian gap change estimate is -0.36 $[-0.96, 0.23]$, which is not statistically different from zero. In the refreshed segment, the 2026 estimate is -0.32 $[-0.54, -0.09]$, indicating less favourable movement from 2025 to 2026 for North Asian students than for the non-North Asian comparison.

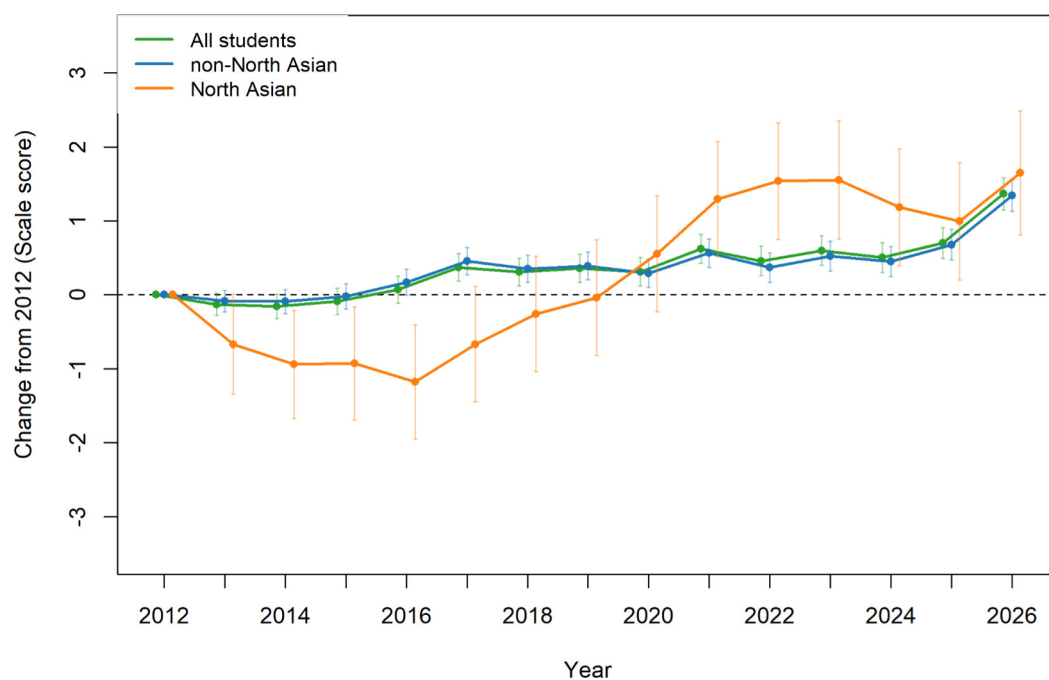


FIGURE 10e. North Asian Reading Vocabulary focal trajectory

Students who identified as North Asian end the observation window in Reading Vocabulary at 1.65 [0.81, 2.48], above baseline and above the estimate for students who did not identify as North Asian of 1.34 [1.13, 1.56]. The line rises strongly from the late 2010s, softens after 2022, and rises again by 2026.

This reinforces that Reading Vocabulary and Reading Comprehension should not be treated as the same literacy outcome. The North Asian Reading Vocabulary trajectory is positive, while the North Asian Reading Comprehension trajectory is below baseline in the legacy segment and only weakly positive in the refreshed segment.

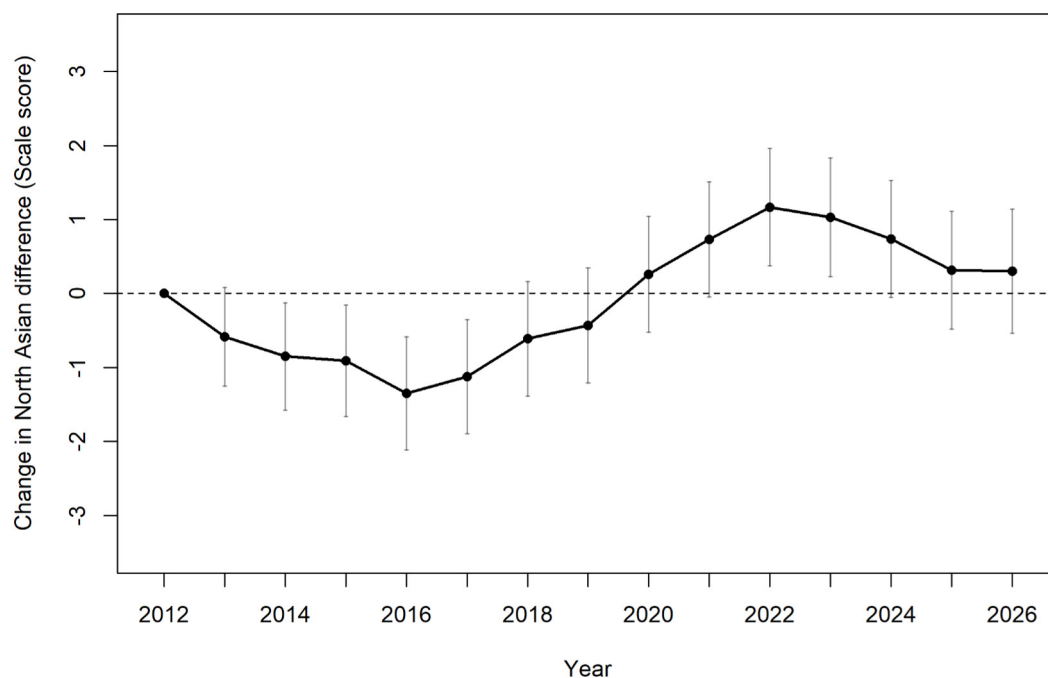


FIGURE 10f. **North Asian relative to non-North Asian common-reference difference-over-time for PAT Reading Vocabulary**

The common-reference Reading Vocabulary output is positive in point estimate terms, but less decisively so than the trajectory plot. The final North Asian relative to non-North Asian common-reference gap change estimate is 0.30 [-0.54, 1.14]. The interval includes zero.

North Asian Reading Vocabulary increased, but the common-reference interval includes zero. The analysis therefore does not establish a clear relative Reading Vocabulary gain over the comparison group.

Taken together, the North Asian results are strongest in Mathematics. The legacy Mathematics trajectory and common-reference gap both move clearly in a favourable direction, and the refreshed Mathematics segment is also positive. Reading Vocabulary is positive but not clearly a relative gain under the common-reference comparison. Reading Comprehension is the exception. It declines in the legacy segment, and the first refreshed movement is less favourable than for the non-North Asian comparison. This profile matters because North Asian students have high absolute Mathematics levels, but their within-series trajectories are not uniformly positive across all three PAT measures.

9.7 South, Southeast, and Other Asian

South, Southeast, and Other Asian results are broadly positive in Mathematics and more mixed in the two reading measures. The legacy Mathematics segment ends slightly above its 2012 starting point, and the refreshed Mathematics segment rises clearly from 2024 to 2026. Reading Comprehension follows the wider legacy decline, while the refreshed segment rises modestly from 2025 to 2026. Reading Vocabulary is positive by 2026, though the common-reference gap estimate remains uncertain.

TABLE 13. **South, Southeast, and Other Asian companion summary**

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final comparison trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	0.43 [0.01, 0.86]	-0.07 [-0.20, 0.05]	0.51 [0.07, 0.95]
Mathematics refreshed	2024 to 2026	2.03 [1.83, 2.24]	1.57 [1.49, 1.66]	0.46 [0.23, 0.69]
Reading Comprehension legacy	2012 to 2024	-1.70 [-2.23, -1.18]	-1.68 [-1.82, -1.53]	-0.03 [-0.58, 0.52]
Reading Comprehension refreshed	2025 to 2026	0.39 [0.21, 0.58]	0.51 [0.44, 0.58]	-0.12 [-0.33, 0.09]
Reading Vocabulary	2012 to 2026	1.02 [0.30, 1.74]	1.43 [1.21, 1.65]	-0.40 [-1.15, 0.34]

Students identified as South, Southeast, and Other Asian end the legacy Mathematics segment above their 2012 baseline, at 0.43 [0.01, 0.86]. The comparison group is close to baseline, at -0.07 [-0.20, 0.05]. The focal interval now sits narrowly above zero, supporting a modest positive legacy result.

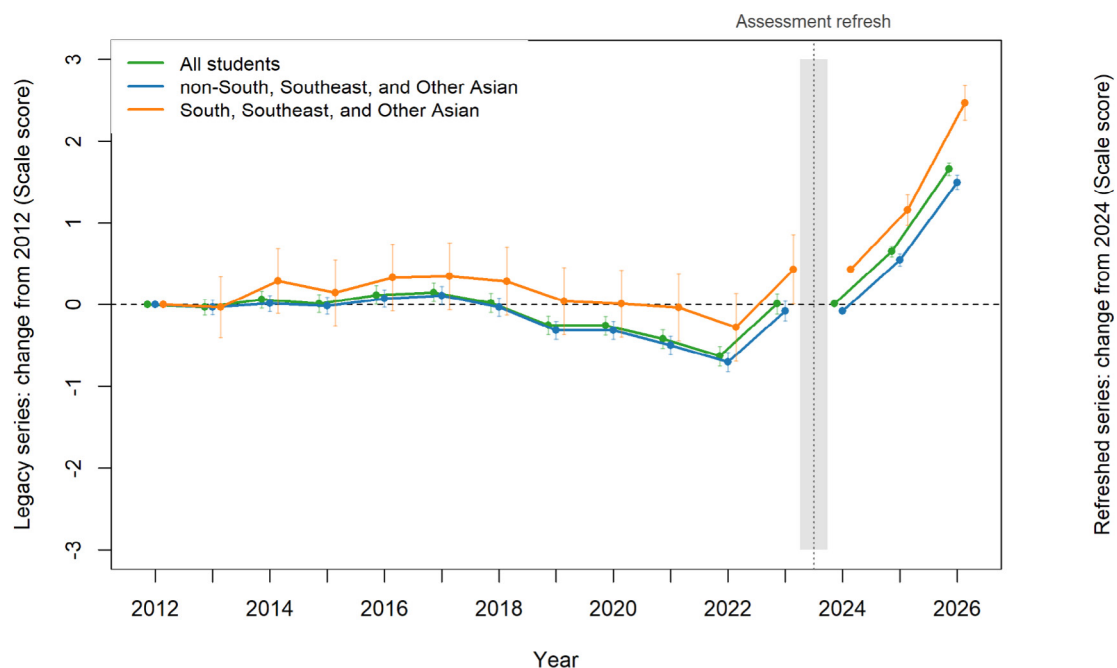


FIGURE 11a. **South, Southeast, and Other Asian Mathematics focal trajectory**

Within the refreshed Mathematics segment, students identified as South, Southeast, and Other Asian rise 2.03 [1.83, 2.24] from the 2024 refreshed baseline by 2026. The comparison group also rises, at 1.57 [1.49, 1.66]. The refreshed pattern is therefore positive for both groups, with the focal movement somewhat larger.

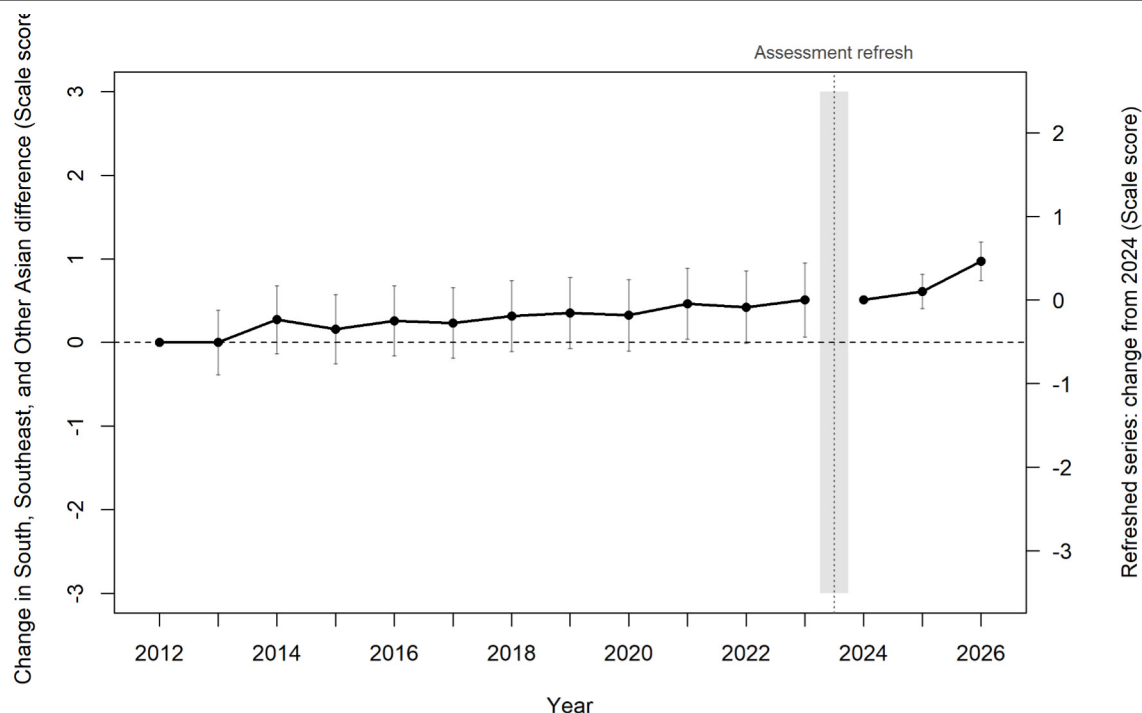


FIGURE 11b. **South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Mathematics**

The common-reference Mathematics output supports a modest favourable relative interpretation. The final legacy South, Southeast, and Other Asian relative to comparison-group change estimate is 0.51 [0.07, 0.95]. In the refreshed segment, the corresponding 2026 estimate is 0.46 [0.23, 0.69]. In both segments, the focal group moves more favourably than the comparison group.

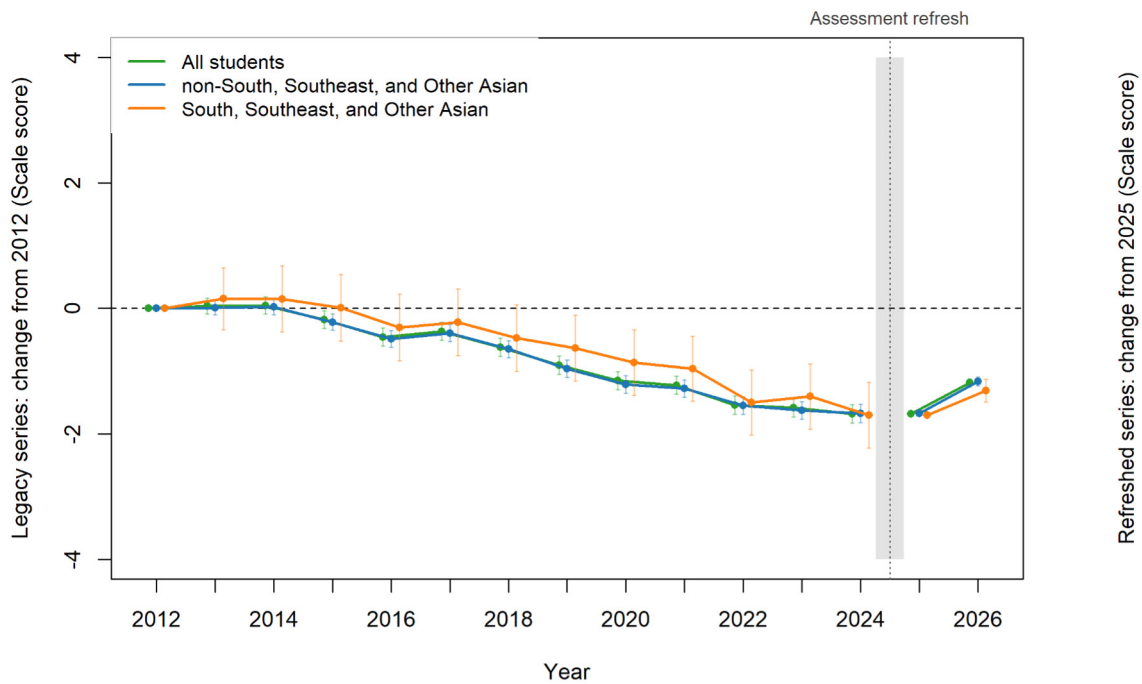


FIGURE 11c. **South, Southeast, and Other Asian Reading Comprehension focal trajectory**

Reading Comprehension tells a more cautious story. Students identified as South, Southeast, and Other Asian end the legacy segment below baseline at -1.70 [-2.23, -1.18]. The comparison group also ends below baseline, at -1.68 [-1.82, -1.53]. The focal line therefore fits the wider Reading Comprehension decline rather than standing apart from it.

Within the refreshed Reading Comprehension segment, the focal group rises 0.39 [0.21, 0.58] from 2025 to 2026. The comparison group also rises, at 0.51 [0.44, 0.58]. This indicates early positive refreshed movement for both groups, with a smaller point estimate for students identified as South, Southeast, and Other Asian.

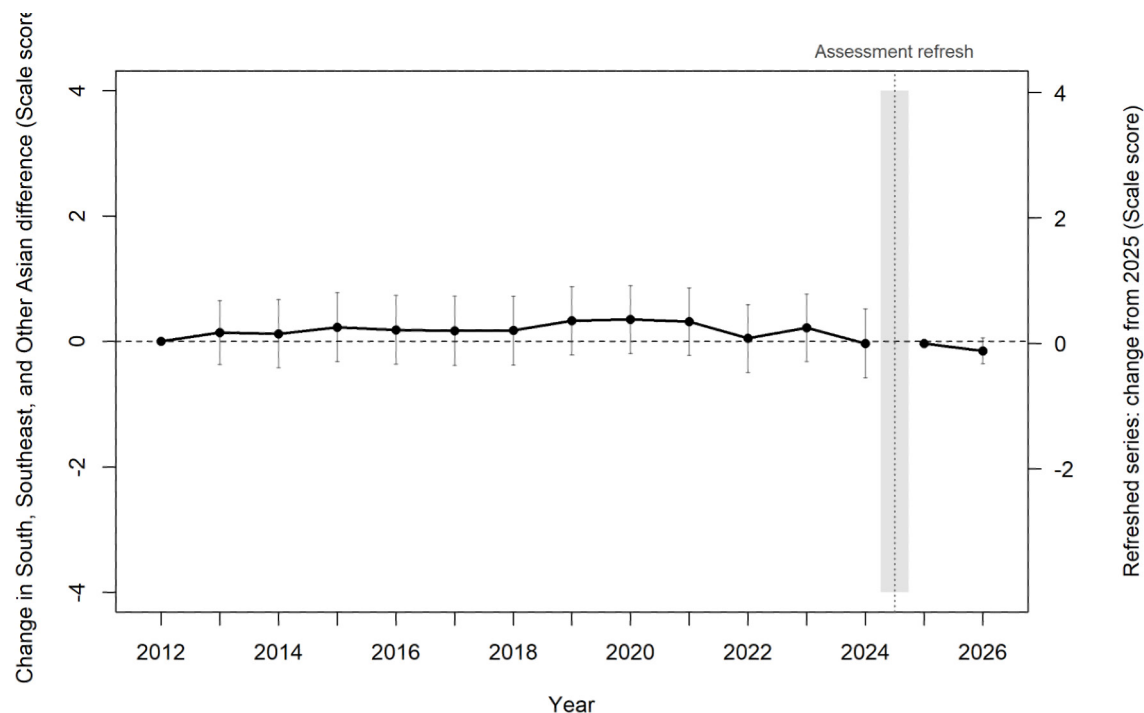


FIGURE 11d. **South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Reading Comprehension**

The common-reference Reading Comprehension output is close to zero. The final legacy change estimate is -0.03 [-0.58, 0.52], and the refreshed 2026 estimate is -0.12 [-0.33, 0.09]. Both intervals include zero, so the gap figure does not support a strong claim of favourable or unfavourable relative movement.

This changes how the trajectory plot should be read. The focal group and comparison group both declined in the legacy series and both improved within the refreshed series. The common-reference output suggests that these movements were not clearly different once both groups are evaluated against the same reference population.

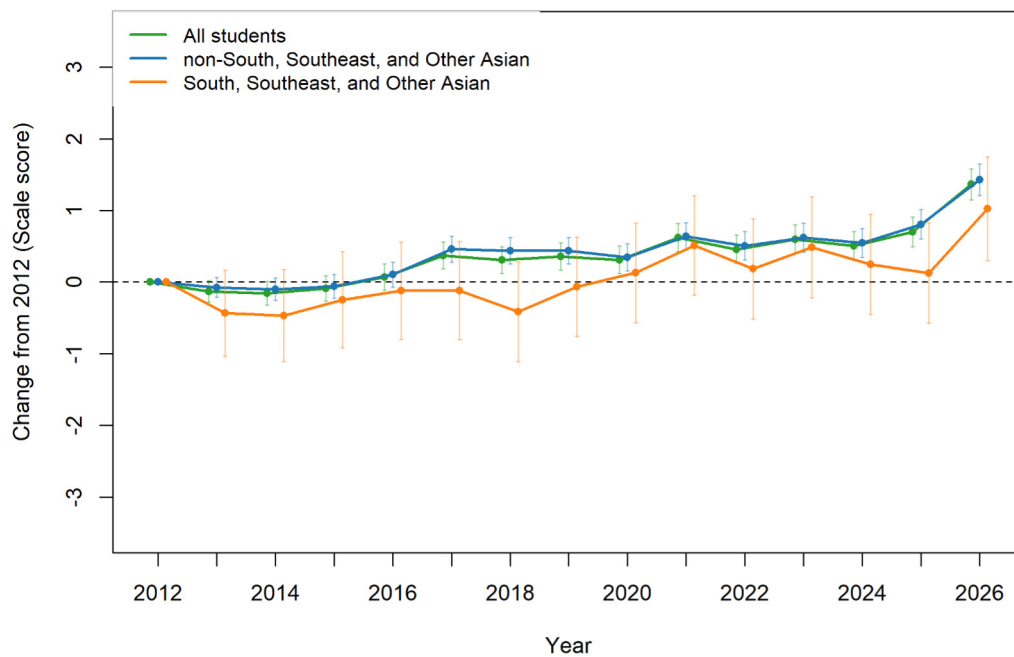


FIGURE 11e. **South, Southeast, and Other Asian Reading Vocabulary focal trajectory**

Students identified as South, Southeast, and Other Asian end above baseline at 1.02 [0.30, 1.74]. The comparison group also ends above baseline, at 1.43 [1.21, 1.65]. The focal Vocabulary trajectory is positive by 2026, but the comparison estimate is higher and the focal line is not a smooth upward series across the full period.

This result fits the broader Reading Vocabulary pattern in the report. Vocabulary movement is more favourable than Reading Comprehension, but the South, Southeast, and Other Asian Vocabulary profile should not be read as a uniform or especially strong gain across the whole observation window.

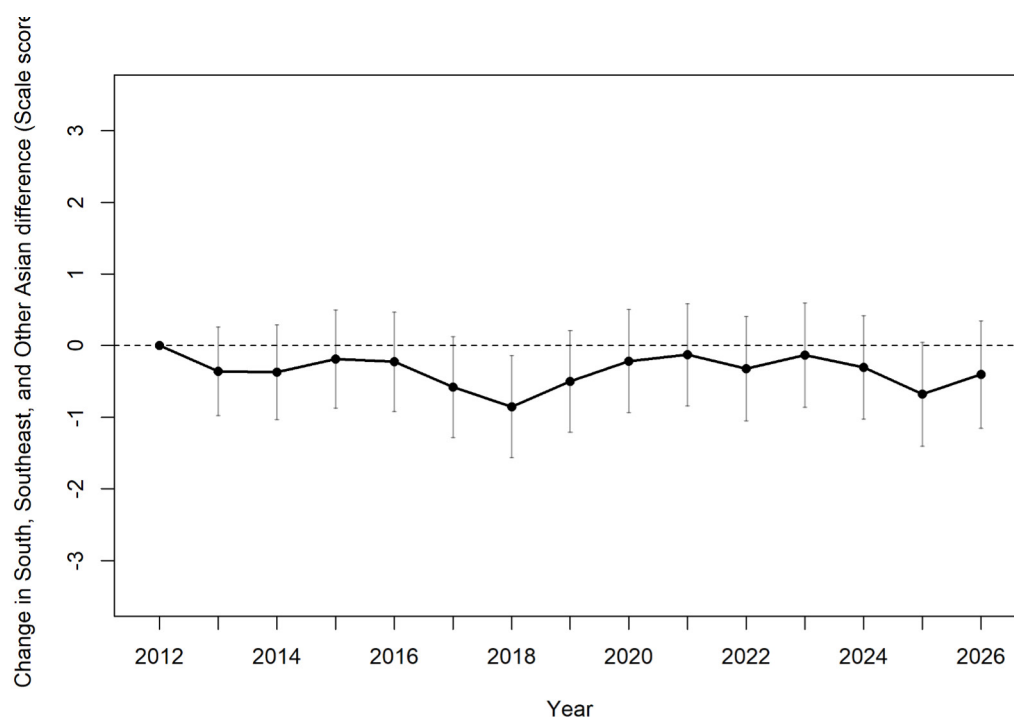


FIGURE 11f. **South, Southeast, and Other Asian relative to not South, Southeast, and Other Asian same-starting-mix difference-over-time for PAT Reading Vocabulary**

The common-reference Vocabulary output is more cautious than the trajectory plot. The final change estimate is -0.40 [-1.15, 0.34]. The interval includes zero, so the analysis does not establish a relative change between the South, Southeast, and Other Asian focal group and the rest of the tested population.

The main message is that students identified as South, Southeast, and Other Asian show favourable Mathematics movement and a less distinctive reading profile. Mathematics shows positive movement in both assessment segments and modest favourable common-reference gap movement. Reading Comprehension follows the broader legacy decline and then rises within the refreshed segment, without clear relative gap change from the rest of students sitting PAT assessments. Reading Vocabulary is positive by 2026, but not clearly more favourable than the comparison group.

9.8 African

This subsection reports students identified as African. In general, annual focal counts are smaller than for the previous ethnic categories, so confidence intervals are often wider and the exact year-to-year changes should therefore be read with correspondingly greater care. The African category is also broad. It includes students with potentially very different migration histories, cultures, language backgrounds, and schooling experiences.

TABLE 14. African companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final non-group trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	-0.69 [-1.51, 0.13]	0.03 [-0.09, 0.15]	-0.72 [-1.53, 0.10]
Mathematics refreshed	2024 to 2026	1.80 [1.39, 2.21]	1.65 [1.57, 1.73]	0.16 [-0.25, 0.57]
Reading Comprehension legacy	2012 to 2024	-3.67 [-4.67, -2.66]	-1.63 [-1.78, -1.48]	-2.04 [-3.03, -1.04]
Reading Comprehension refreshed	2025 to 2026	0.69 [0.32, 1.06]	0.49 [0.42, 0.55]	0.20 [-0.17, 0.57]
Reading Vocabulary	2012 to 2026	1.06 [-0.28, 2.41]	1.37 [1.16, 1.59]	-0.31 [-1.65, 1.03]

The Mathematics trajectory is mixed across assessment segments. Students identified as African end the legacy segment at -0.69 [-1.51, 0.13], while the non-African comparison ends at 0.03 [-0.09, 0.15]. The focal interval includes zero. In the refreshed segment, both groups rise, with students identified as African at 1.80 [1.39, 2.21] and the non-African comparison at 1.65 [1.57, 1.73].

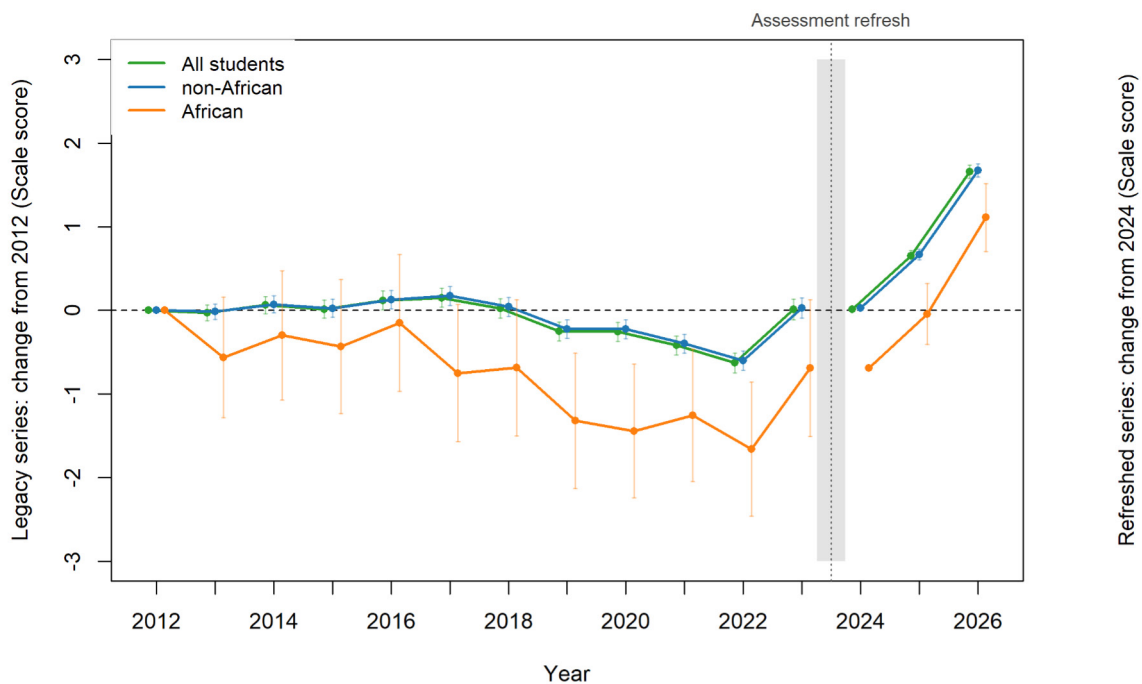


FIGURE 12a. African Mathematics focal trajectory

In general, the legacy trajectory line for African students is below the non-African line for much of the observation window, but the confidence intervals overlap at the endpoint. The refreshed segment is positive for both lines, with the African point estimate slightly higher by 2026.

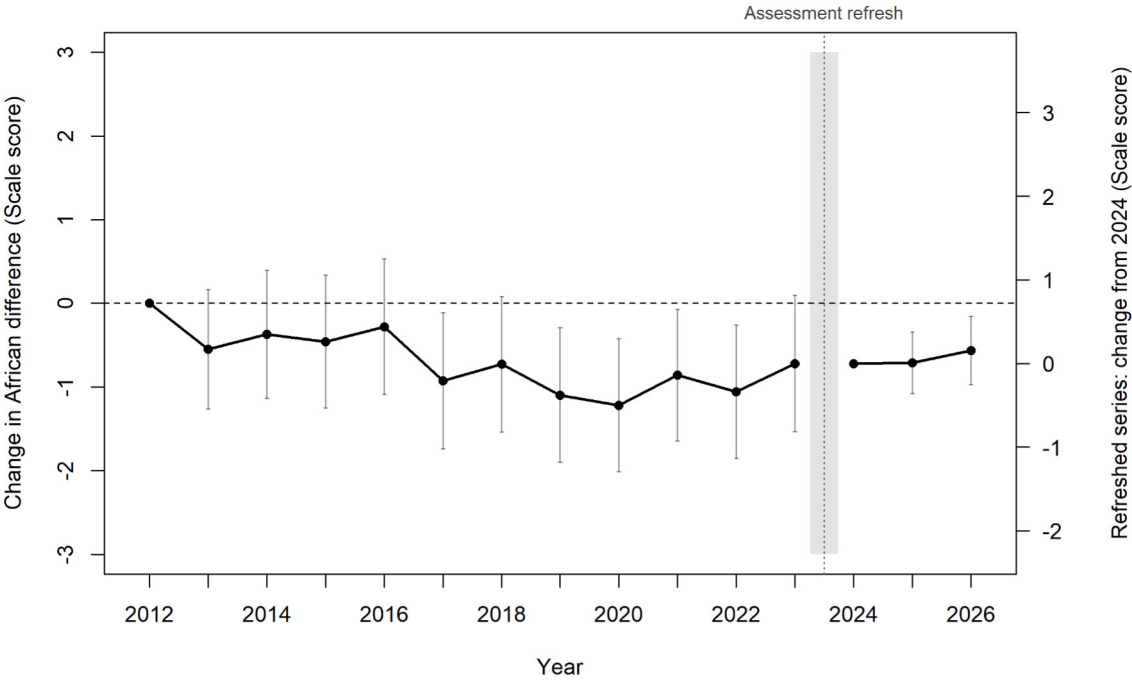


FIGURE 12b. African relative to not African common-reference difference-over-time for PAT Mathematics

The final legacy common-reference Mathematics estimate is -0.72 [-1.53, 0.10]. This point estimate is negative, but the interval includes zero at the endpoint. The refreshed common-reference estimate is 0.16 [-0.25, 0.57], so Mathematics does not show a clear relative gap movement in the refreshed segment.

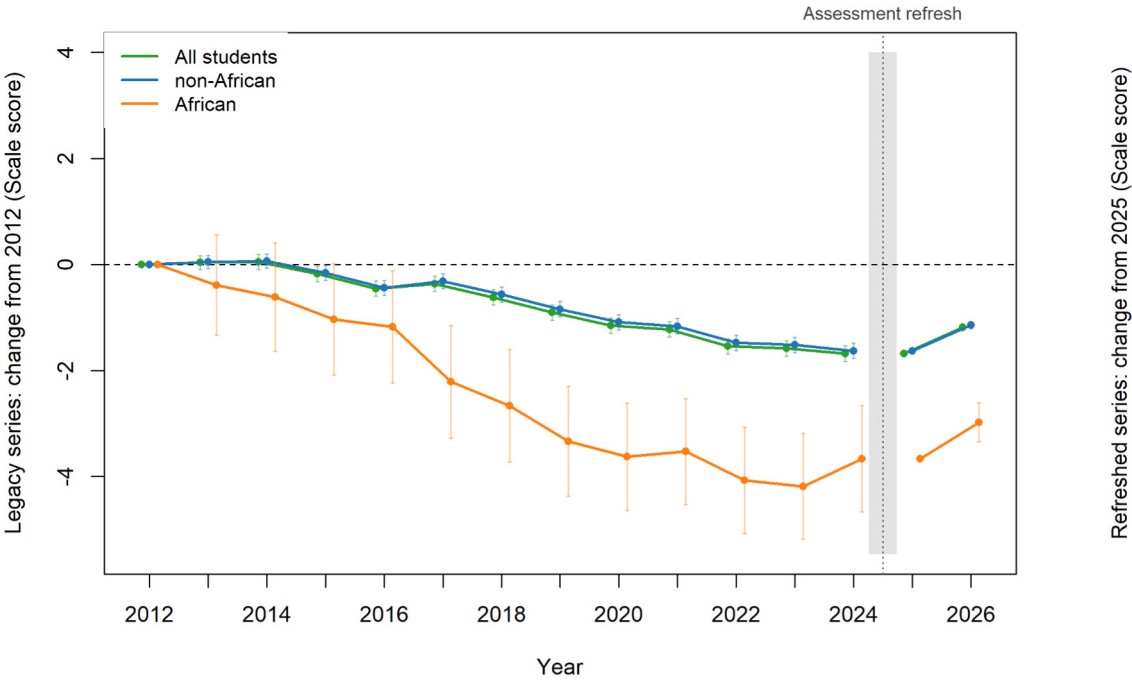


FIGURE 12c. African Reading Comprehension focal trajectory

In contrast, PAT Reading Comprehension provides a much clearer legacy signal. Students identified as African end the legacy segment at -3.67 [-4.67, -2.66], while the non-African comparison ends at -1.63 [-1.78, -1.48]. Both trajectories are below baseline, but the African trajectory falls much further from its own 2012 starting point. The confidence intervals do not overlap.

The African legacy result is a more negative instance of the wider Reading Comprehension decline.

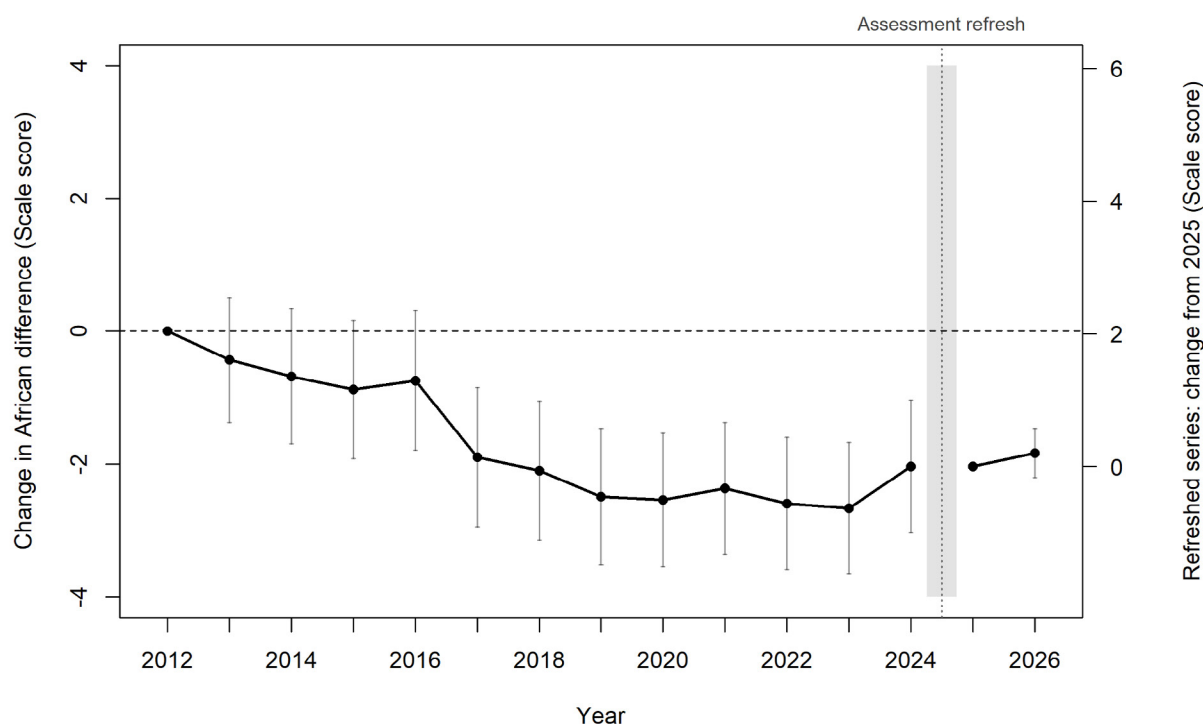


FIGURE 12d. **African relative to non-African common-reference difference-over-time for PAT Reading Comprehension**

The common-reference Reading Comprehension figure strengthens the picture in the trajectory figure above. The final legacy African relative to non-African change estimate is -2.04 [-3.03, -1.04]. This means that, under a common-reference comparison, the African relative position in Reading Comprehension moved about two scale-score points in a less favourable direction from its 2012 value.

The refreshed Reading Comprehension segment is more positive. Students identified as African rise 0.69 [0.32, 1.06] from 2025 to 2026, while the non-African comparison rises 0.49 [0.42, 0.55]. The refreshed common-reference estimate is 0.20 [-0.17, 0.57], with an interval that includes zero.

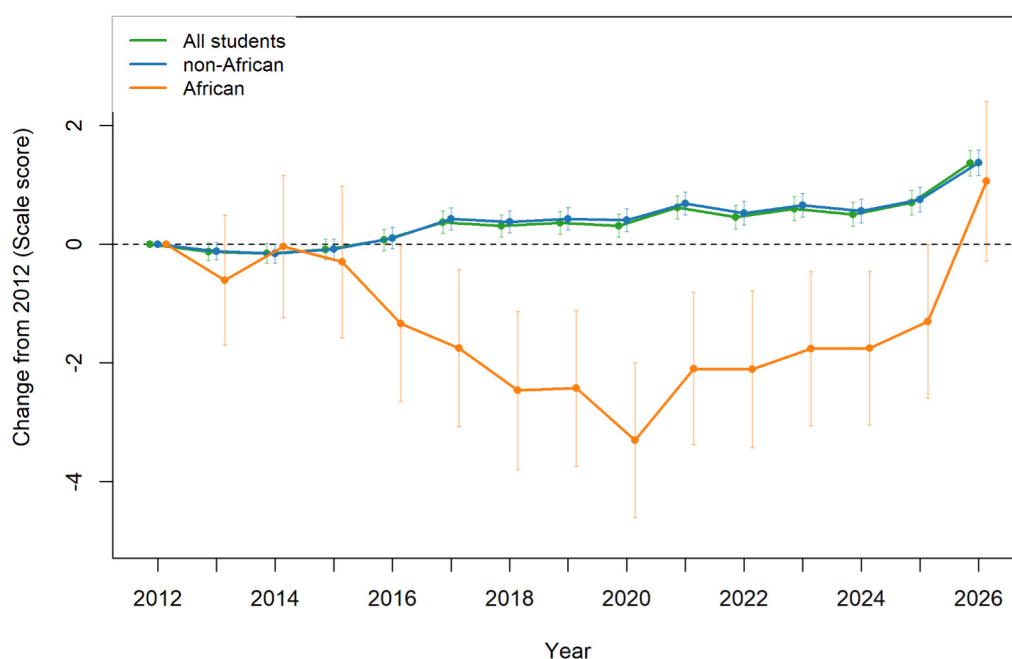


FIGURE 12e. **African Reading Vocabulary focal trajectory**

Reading Vocabulary is more mixed. Students identified as African end above baseline at 1.06 [-0.28, 2.41], although this focal endpoint is imprecise and includes zero. The non-African comparison also ends above baseline, at 1.37 [1.16, 1.59].

The trajectory plot shows a pronounced dip around 2020 before recovery by 2026. Unlike Reading Comprehension, the 2026 Vocabulary evidence does not show a clear negative relative signal for students identified as African.

This matters because Reading Vocabulary has generally been the more positive reading-related series in the report. The African profile also ends in a positive direction, but with wider uncertainty than for the non-African comparison.

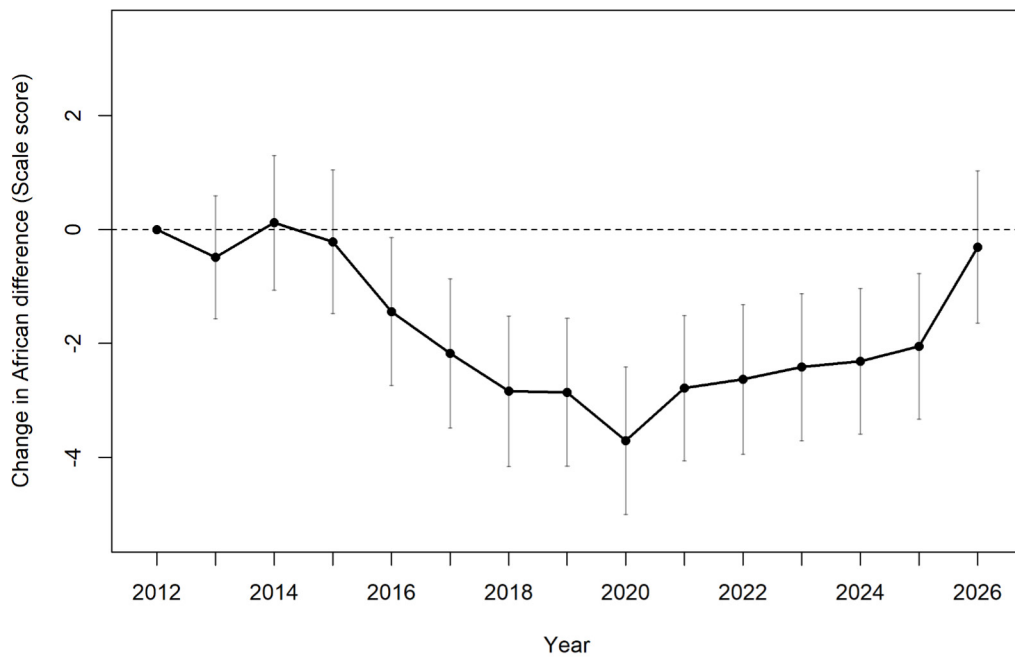


FIGURE 12f. **African relative to non-African common-reference difference-over-time for PAT Reading Vocabulary**

The common-reference Vocabulary output is also cautious. The final African relative to non-African change estimate is -0.31 [-1.65, 1.03]. The interval includes zero.

Overall, the clearest African result is the more negative legacy Reading Comprehension trajectory and common-reference difference. Mathematics is less decisive in the legacy segment and positive in the refreshed segment; Vocabulary is positive in point-estimate terms by 2026 but remains imprecise.

9.9 Middle Eastern

This subsection reports students identified as Middle Eastern. Annual counts are smaller than for the three largest focal groups, especially in Reading Vocabulary, as shown in Table 7. The estimates are therefore less precise. The PAT outputs show distinctive positive movement in Mathematics and Reading Vocabulary and a Reading Comprehension trajectory that is less negative than the non-Middle Eastern comparison in the legacy segment.

TABLE 15. Middle Eastern companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final non-group trajectory estimate [95% CI]	Final common-reference gap change estimate [95% CI]
Mathematics legacy	2012 to 2023	1.89 [0.79, 2.99]	-0.01 [-0.14, 0.11]	1.90 [0.80, 3.00]
Mathematics refreshed	2024 to 2026	1.84 [1.28, 2.40]	1.65 [1.57, 1.73]	0.19 [-0.37, 0.76]
Reading Comprehension legacy	2012 to 2024	-0.45 [-1.83, 0.92]	-1.70 [-1.84, -1.55]	1.24 [-0.13, 2.61]
Reading Comprehension refreshed	2025 to 2026	0.46 [-0.04, 0.96]	0.49 [0.43, 0.56]	-0.03 [-0.53, 0.47]
Reading Vocabulary	2012 to 2026	3.49 [1.71, 5.26]	1.34 [1.12, 1.55]	2.15 [0.39, 3.92]

The Mathematics trajectory is sharply positive. Students identified as Middle Eastern end the legacy segment at 1.89 [0.79, 2.99], while the non-Middle Eastern comparison ends near baseline at -0.01 [-0.14, 0.11]. The refreshed segment is also positive, with Middle Eastern students rising 1.84 [1.28, 2.40] from 2024 to 2026.

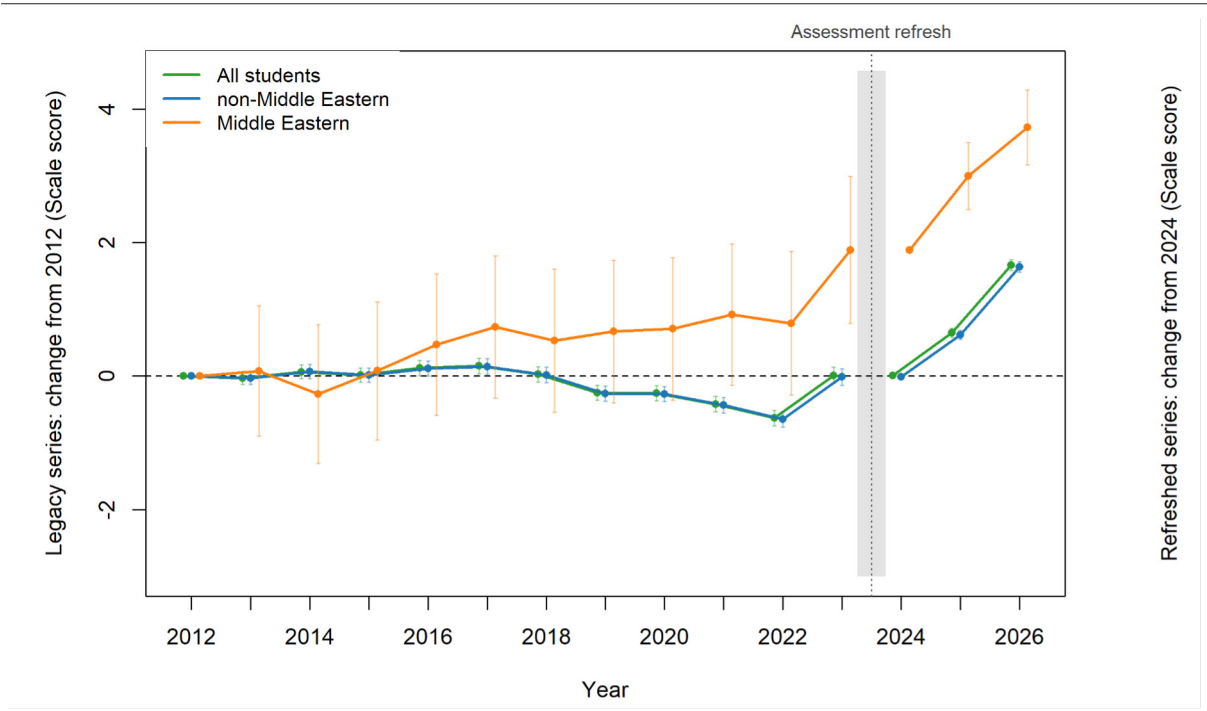


FIGURE 13a. Middle Eastern Mathematics one-vs-rest focal trajectory

The shape of the line matters. The focal Mathematics trajectory begins near baseline or mildly below it, then rises into a clearly positive later legacy position. The refreshed segment continues in a positive direction.

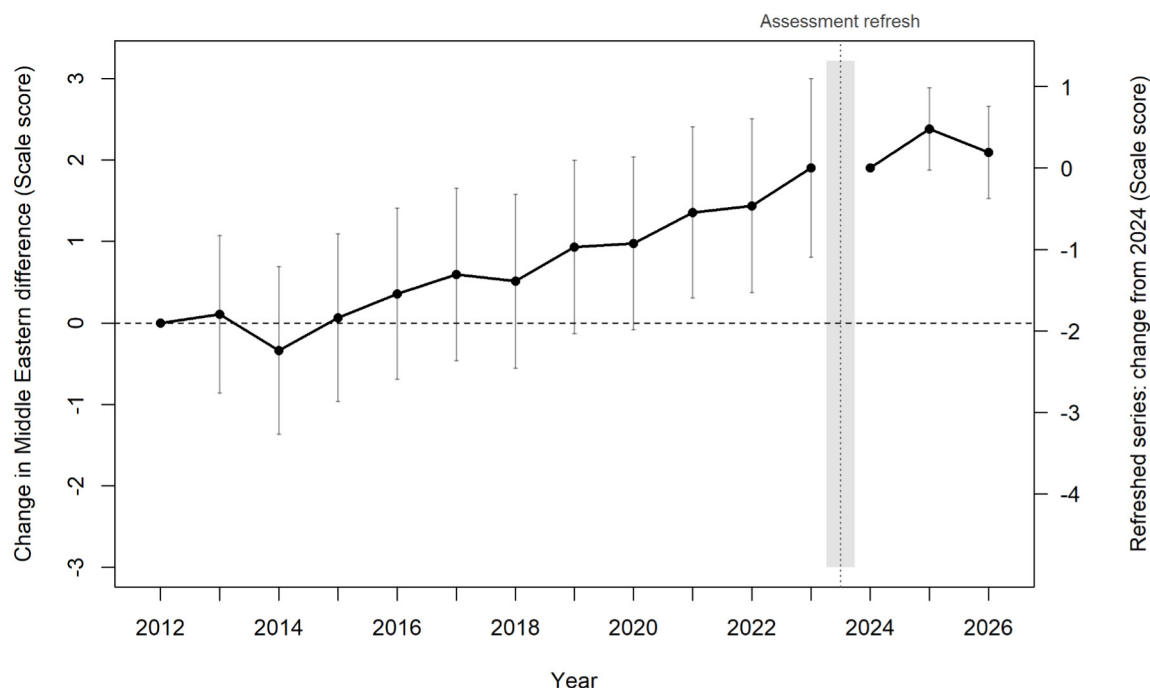


FIGURE 13b. Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Mathematics

The common-reference Mathematics output reinforces the positive relative interpretation in the legacy segment. The final legacy Middle Eastern relative to non-Middle Eastern change estimate is 1.90 [0.80, 3.00]. This means that, under a common-reference comparison, the estimate for the Middle Eastern relative position in Mathematics moved almost two scale-score points in a more favourable direction from its 2012 value. In the refreshed segment, the final relative estimate is 0.19 [-0.37, 0.76], which does not establish a further difference.

This result is less precise than the large group profiles because the annual number of Middle Eastern children who sat PAT Mathematics was smaller, but the legacy signal is still strong, with confidence intervals that do not include zero in the final legacy year.

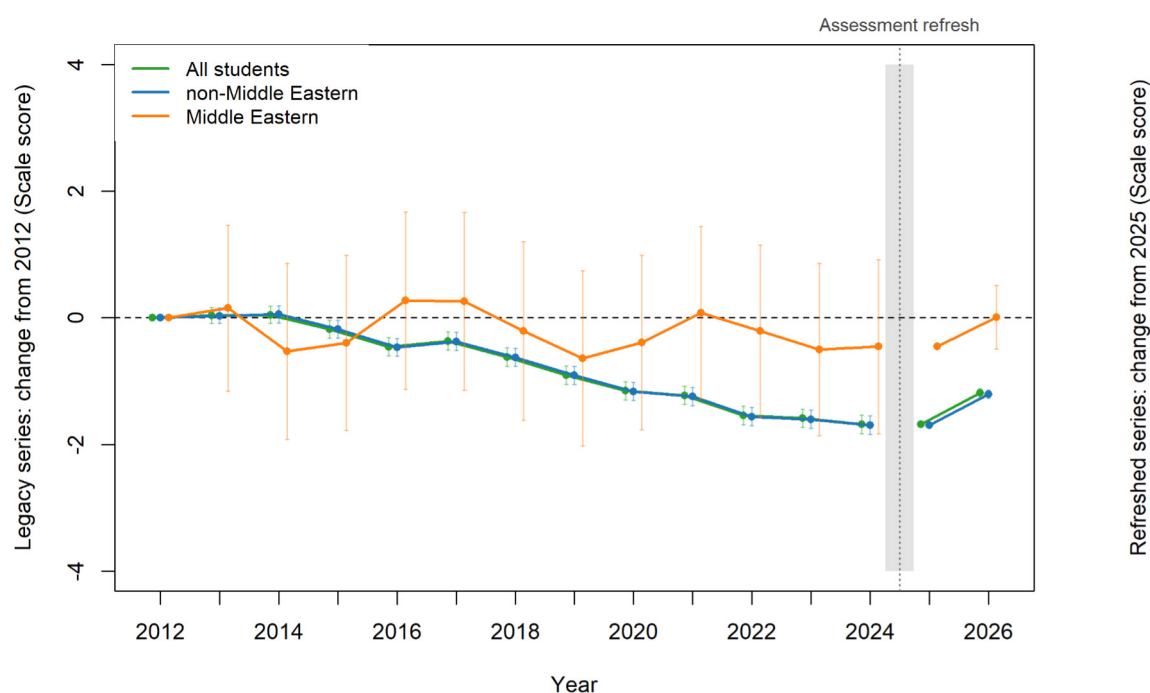


FIGURE 13c. Middle Eastern Reading Comprehension one-vs-rest focal trajectory

Reading Comprehension gives a different kind of result. Students identified as Middle Eastern end the legacy segment at -0.45 $[-1.83, 0.92]$, while the non-Middle Eastern comparison ends below baseline at -1.70 $[-1.84, -1.55]$. In the refreshed segment, Middle Eastern students rise 0.46 $[-0.04, 0.96]$ from 2025 to 2026, while the non-Middle Eastern comparison rises 0.49 $[0.43, 0.56]$.

This is distinctive since Reading Comprehension is the most broadly negative assessment series in this report. Most focal groups end the legacy segment clearly below their early reference point. Middle Eastern Reading Comprehension, in contrast, ends closer to zero in point estimate terms. The wide confidence interval means that the estimate is uncertain. The legacy estimate overlaps with zero and also overlaps with the non-Middle Eastern comparison.

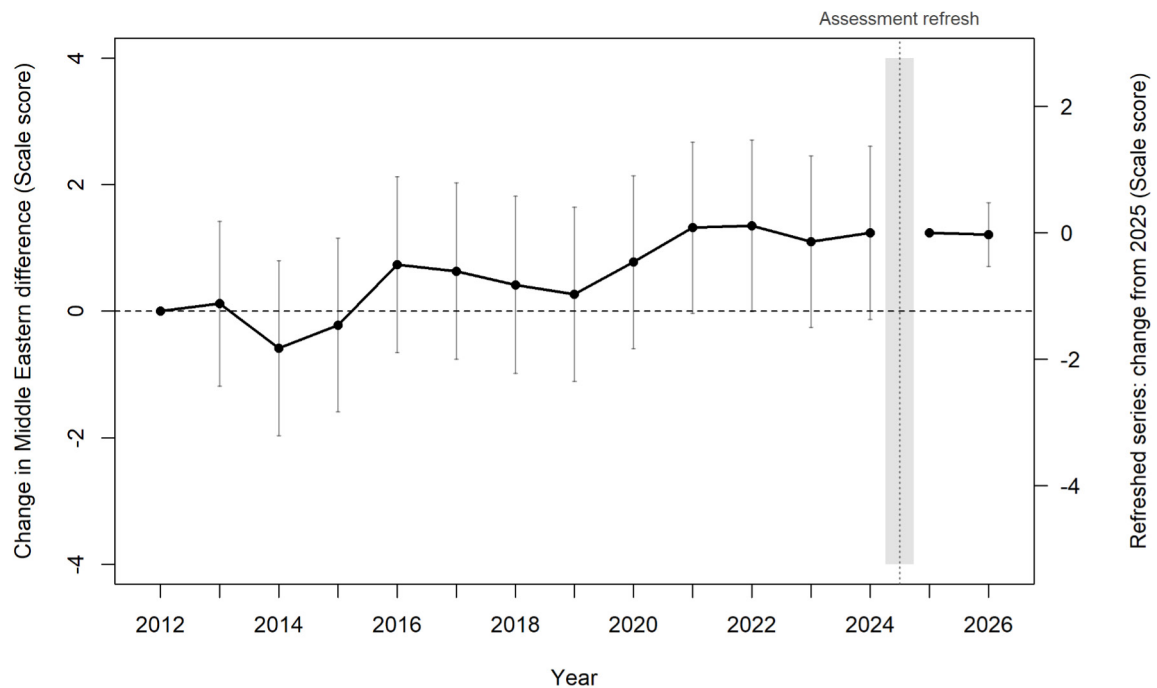


FIGURE 13d. Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Reading Comprehension

The common-reference Reading Comprehension output is positive in point estimate terms in the legacy segment, though yearly confidence intervals typically include zero. The final legacy change estimate is 1.24 [-0.13, 2.61].

The common-reference point estimate is positive in the legacy segment, but its confidence interval includes zero. The analysis therefore does not establish a favourable relative change. In the refreshed segment, the 2026 estimate is -0.03 [-0.53, 0.47], also providing no clear relative difference.

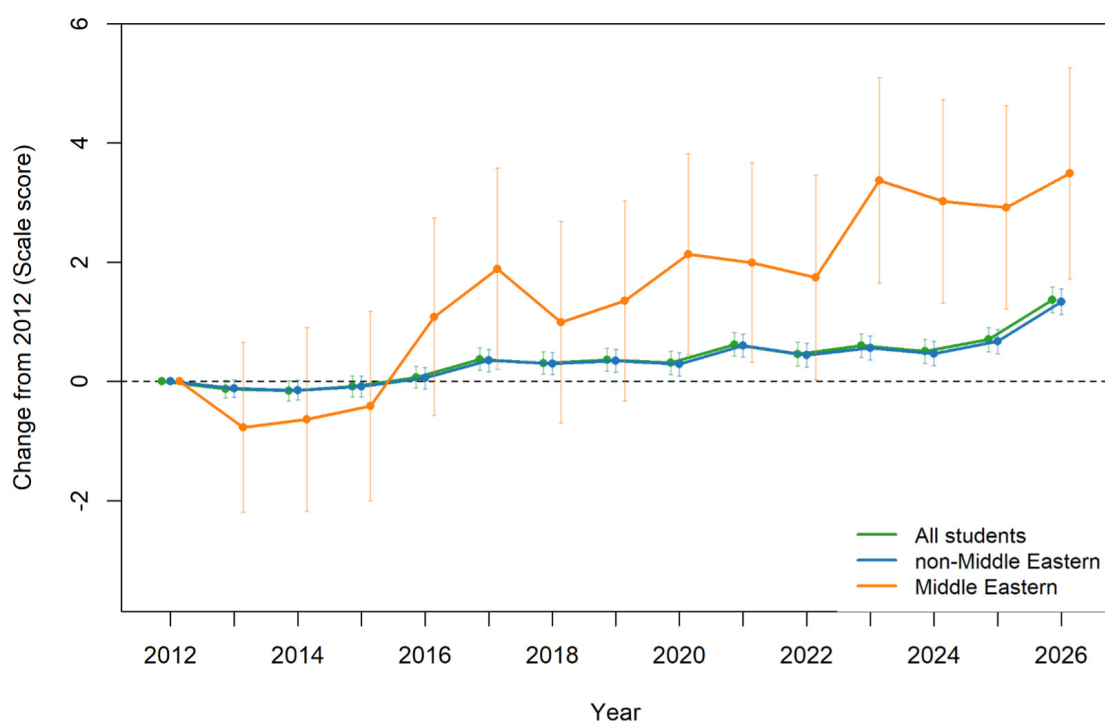


FIGURE 13e. **Middle Eastern Reading Vocabulary focal trajectory**

Reading Vocabulary is strongly positive. Students identified as Middle Eastern end at 3.49 [1.71, 5.26], while the non-Middle Eastern comparison ends at 1.34 [1.12, 1.55]. In the trajectory plot, the focal line sits well above the other lines, especially from the early 2020s onward.

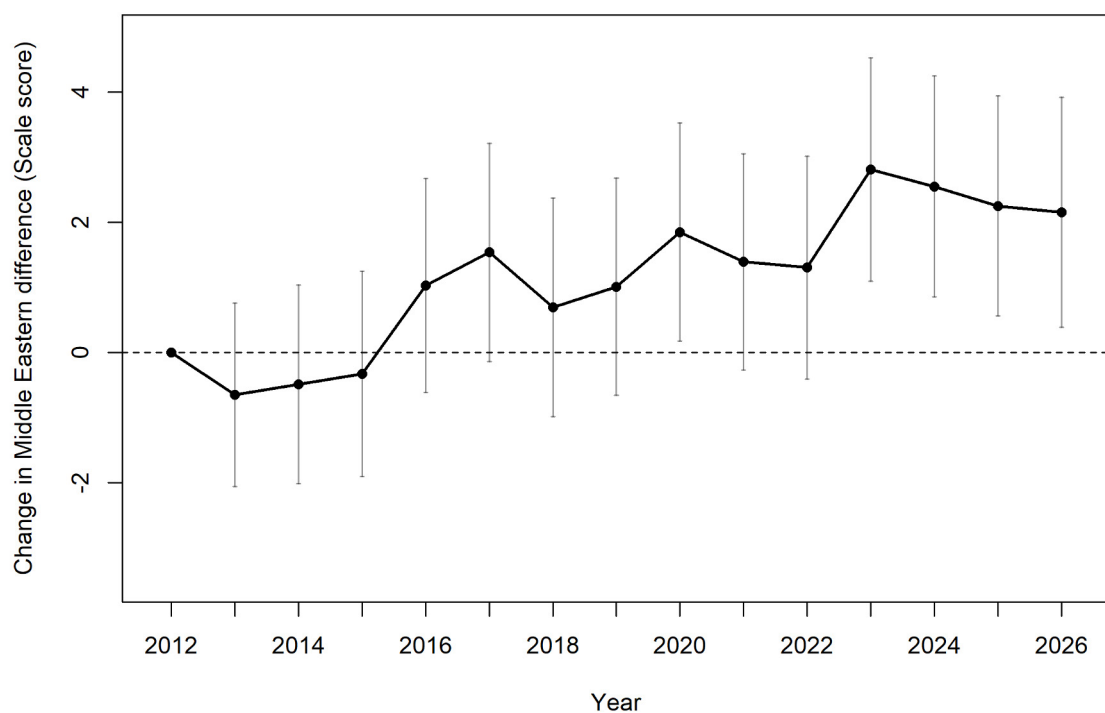


FIGURE 13f. **Middle Eastern relative to non-Middle Eastern common-reference difference-over-time for PAT Reading Vocabulary**

The common-reference Vocabulary output also supports the positive relative interpretation. The final Middle Eastern relative to non-Middle Eastern change estimate is 2.15 [0.39, 3.92]. Under the common-reference comparison, the Middle Eastern relative position in Vocabulary therefore moved more than two scale-score points in a favourable direction from its 2012 value.

Overall, students identified as Middle Eastern show strong positive movement in PAT Mathematics and PAT Reading Vocabulary, with common-reference outputs supporting both relative findings in the legacy Mathematics and continuous Vocabulary series. Reading Comprehension ended less negatively than the non-Middle Eastern comparison in the legacy segment, though wide confidence intervals make this conclusion less certain. The refreshed Mathematics and Reading Comprehension results are positive within their own segments but do not show clear relative separation by 2026.

9.10 Latin American

This profile reports students identified with the Latin American focal indicator. Table 7 shows that this group has the smallest annual counts in the ethnicity chapter, especially in Reading Vocabulary. The estimates can still identify signals in the PAT frame, but their exact year-to-year shape is less precise than for larger focal groups.

The Latin American category is also broad. It may include students with different migration histories, home language backgrounds, schooling experiences, and lengths of time in Aotearoa New Zealand. The interpretation therefore stays close to the modelled PAT outputs and treats the results as visible signals in the PAT frame rather than settled explanations of a uniform group.

TABLE 16. Latin American companion summary

Measure and series	Years	Final focal trajectory estimate [95% CI]	Final not-group trajectory estimate [95% CI]	Final same-starting-mix change estimate [95% CI]
Mathematics legacy	2012 to 2023	1.60 [0.03, 3.16]	-0.01 [-0.13, 0.11]	1.61 [0.04, 3.17]
Mathematics refreshed	2024 to 2026	1.95 [1.34, 2.56]	1.65 [1.57, 1.73]	0.31 [-0.31, 0.92]
Reading Comprehension legacy	2012 to 2024	-2.09 [-4.07, -0.11]	-1.68 [-1.82, -1.53]	-0.41 [-2.39, 1.56]
Reading Comprehension refreshed	2025 to 2026	0.32 [-0.21, 0.86]	0.49 [0.43, 0.56]	-0.17 [-0.71, 0.37]
Reading Vocabulary	2012 to 2026	4.66 [2.20, 7.13]	1.33 [1.11, 1.54]	3.34 [0.88, 5.79]

The Mathematics trajectory remains strongly positive. Students identified as Latin American end the legacy Mathematics segment at 1.60 [0.03, 3.16], while the not Latin American comparison is close to baseline at -0.01 [-0.13, 0.11]. In the refreshed segment, the focal line rises to 1.95 [1.34, 2.56] from the 2024 refreshed baseline, while the not Latin American comparison rises to 1.65 [1.57, 1.73].

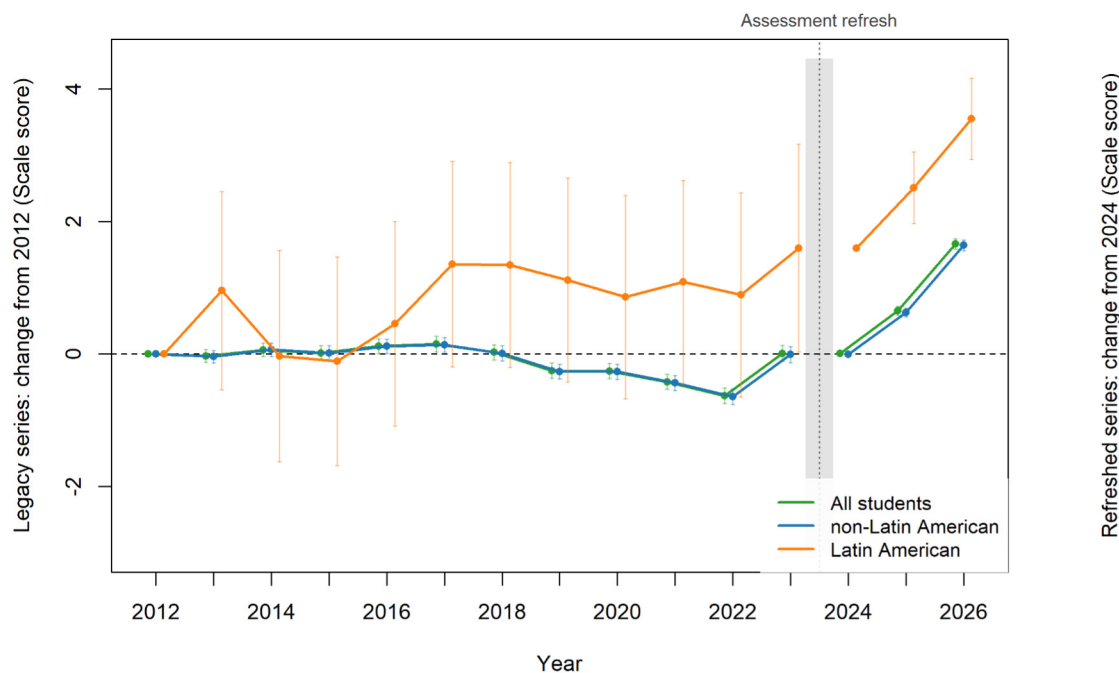


FIGURE 14a. Latin American Mathematics one-vs-rest focal trajectory

The line rises across the later legacy years and again in the refreshed segment, so the endpoint is not just a small final-year separation. However, the wide confidence intervals here are important and the results in some years are only based on a few hundred students.

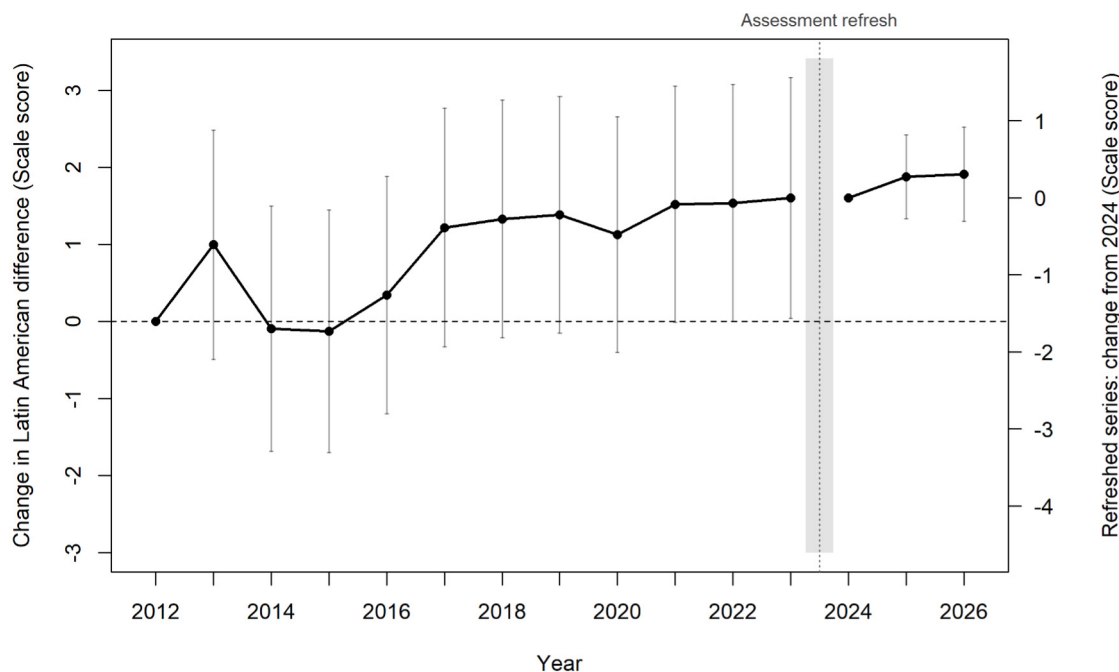


FIGURE 14b. Latin American Mathematics focal-vs-not-group gap change

The final same-starting-mix mathematics legacy estimate is 1.61 [0.04, 3.17] and supports the story told by the trajectory plot. This means that, under the matched starting mix comparison, the Latin American relative position in Mathematics moved in a favourable direction from its 2012 value. In the refreshed segment, the same-starting-mix change is smaller and uncertain at 0.31 [-0.31, 0.92].

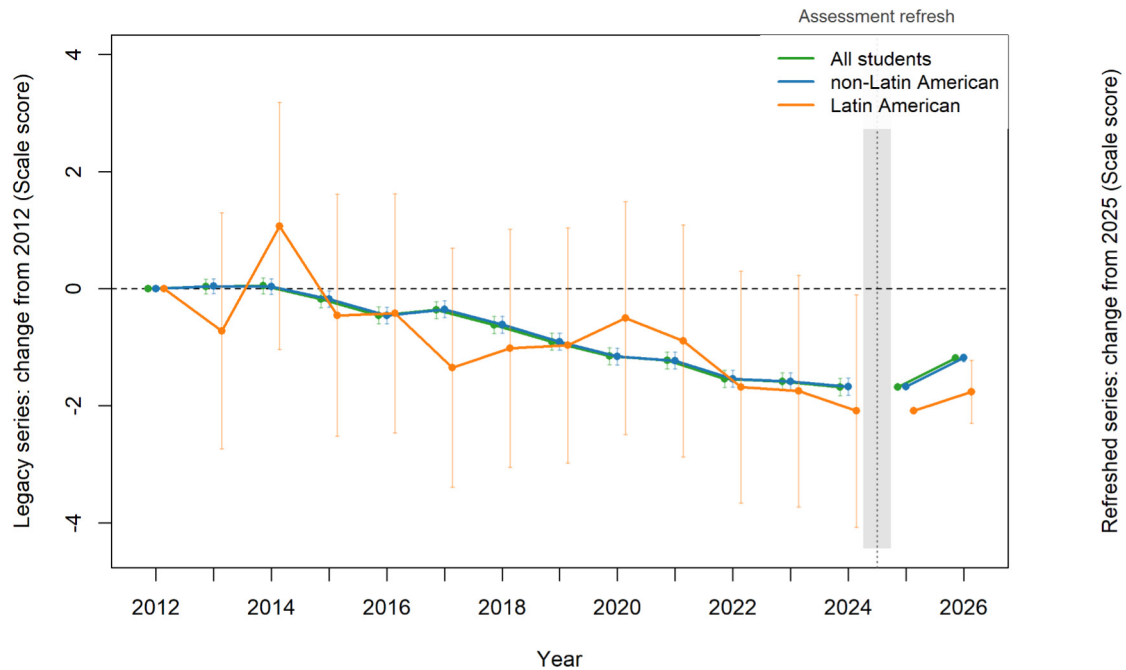


FIGURE 14c. **Latin American Reading Comprehension one-vs-rest focal trajectory**

The trajectory of PAT Reading Comprehension scores for Latin American students is not distinguishable from that of non-Latin American students once the wide confidence intervals are taken into account. Students identified as Latin American end the legacy segment below baseline at -2.09 [-4.07, -0.11], while the not Latin American comparison ends at -1.68 [-1.82, -1.53]. The focal point estimate is more negative than the comparison point estimate, but the focal interval is much wider.

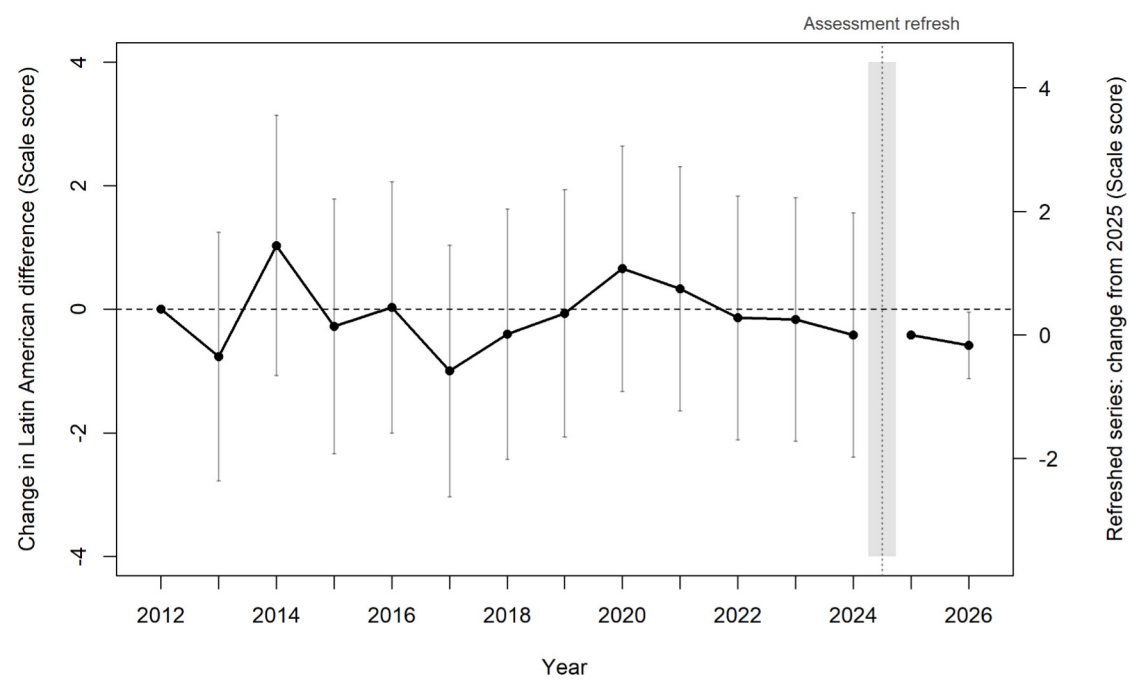


FIGURE 14d. Latin American Reading Comprehension focal-vs-not-group gap change

The final legacy Latin American relative to not Latin American change estimate is -0.41 [-2.39, 1.56]. The final refreshed estimate is -0.17 [-0.71, 0.37]. Both intervals include zero. The estimates are too imprecise to establish that the Latin American Reading Comprehension trajectory differed from the rest of the tested population.

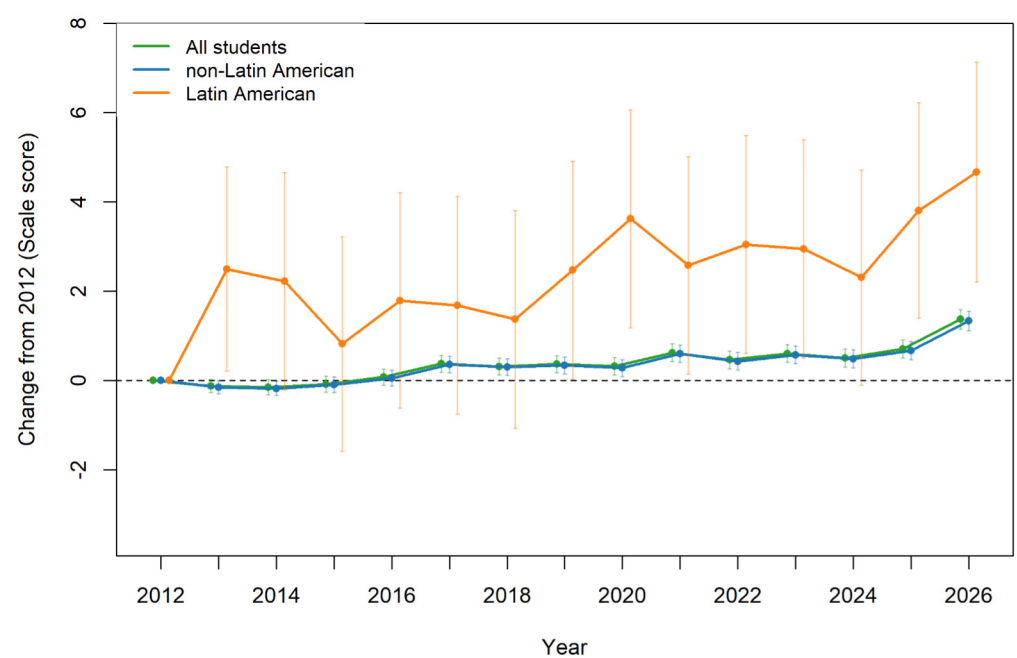


FIGURE 14e. Latin American Reading Vocabulary one-vs-rest focal trajectory

Reading Vocabulary remains at a strongly positive point estimate. Students identified as Latin American end at 4.66 [2.20, 7.13], while the not Latin American comparison ends at 1.33 [1.11, 1.54].

This is one of the largest positive Vocabulary endpoints in the ethnicity chapter, but it is also the result where caution is most important. Few Latin American students took the PAT Reading Vocabulary, and some annual cells are small.

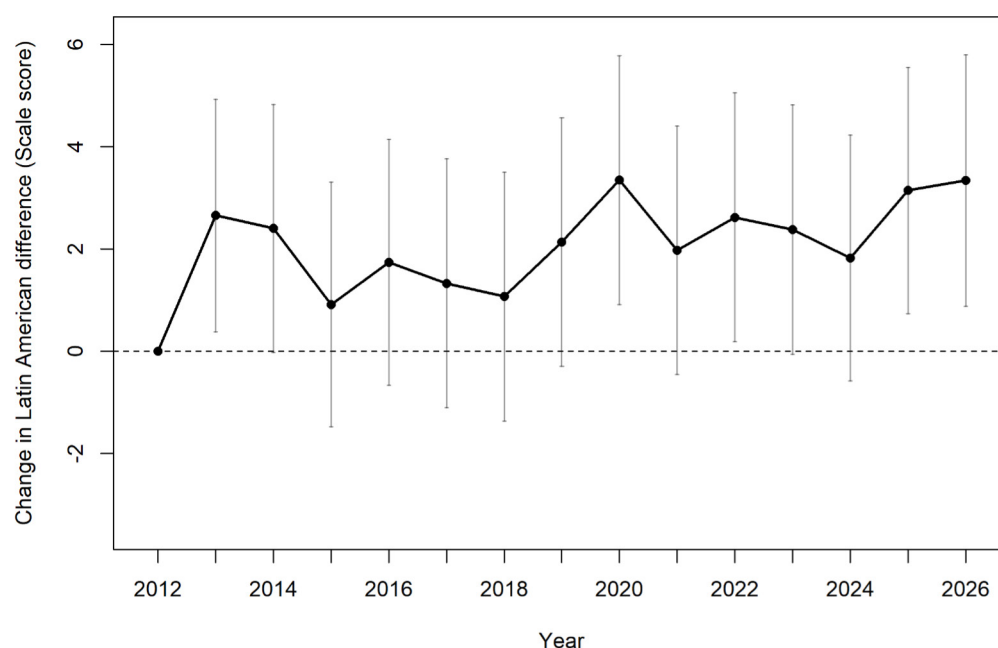


FIGURE 14f. **Latin American Reading Vocabulary focal-vs-not-group gap change**

The same-starting-mix Vocabulary output also supports the positive relative interpretation, though less precisely than the trajectory plot. The final estimate is 3.34 [0.88, 5.79]. Under this comparison, the Latin American relative position in Vocabulary moved in a favourable direction from its 2012 value.

Overall, the Latin American outputs show strong positive signals in Mathematics and Reading Vocabulary, with same-starting-mix results supporting the legacy Mathematics result and the continuous Vocabulary result. Reading Comprehension remains less clear. These results should be read with caution because wide confidence intervals from small annual cell sizes make the exact trajectory uncertain.

9.11 Ethnicity synthesis

The ethnicity results show different trajectories across the three PAT measures. They also show that change from a group's own baseline over time and adjusted achievement differences between groups are separate equity questions. The first asks whether students identified with each group are scoring higher or lower than students identified with that group in the earlier reference period. The second asks how the group's adjusted score level compares with those of other groups after accounting for the other characteristics included in the model. A group can improve relative to its own past while remaining below another group or decline from its own earlier position while remaining above another group. Differences can also narrow because a group with higher adjusted scores has declined, rather than because a group with lower adjusted scores has improved. Such a decline remains a poor outcome

and should generally not be interpreted as equity progress. Improvement within groups and narrowing differences between groups are both important, but neither guarantees the other.

Legacy Reading Comprehension declined for almost every focal group. Pacific students declined less than the comparison group, while African students showed a more negative trajectory, but the dominant result is broad decline across focal groups. Every focal group then moved positively from 2025 to 2026 within the refreshed series.

The Mathematics assessments show a different pattern. In the legacy segment, positive relative movement is strongest for North Asian, South, Southeast, and Other Asian, Middle Eastern, and Latin American students. Māori and European/Pākehā students show less favourable movement, Pacific students end close to baseline, and African estimates are less certain. The refreshed Mathematics segment is positive for every focal group by 2026. However, the absolute plots and common-reference outputs show that improvement did not imply convergence between groups. North Asian Mathematics scores remain particularly separated from other focal groups, while Māori and European/Pākehā profiles show less favourable relative movement than most of the other ethnic group categories.

Most focal groups that declined in legacy Reading Comprehension were above their earlier Vocabulary reference by 2026. The data establish this cross-measure divergence but do not identify its causes.

Across focal groups, within-series improvement did not necessarily coincide with convergence in adjusted achievement levels.

10. Regional patterns

Examining regional patterns provides a spatial check on the national results. The analysis presented in this section explores whether the national patterns were geographically widespread or concentrated in particular parts of the country. Each region is compared with its own within-series reference, so the estimates shown below show change within regions and cannot be used to rank regions.

TABLE 17. Terminal-year regional within-series heatmap values

Region	Mathematics 2023 legacy	Mathematics 2026 refreshed	Reading Comprehension 2024 legacy	Reading Comprehension 2026 refreshed	Reading Vocabulary 2026
All Aotearoa New Zealand	-0.07	1.65	-1.72	0.50	1.53
Bay of Plenty Waairiki	1.16	2.31	-0.12	0.53	2.70
Canterbury Chatham Islands	-0.13	1.21	-1.81	0.30	1.26
Hawke's Bay Tairāwhiti	-1.40	1.89	-2.24	0.54	1.65
Nelson Marlborough West Coast	0.29	1.48	-1.60	0.60	1.50
Otago Southland	-0.81	1.03	-2.08	0.65	1.87
Tai Tokerau	-0.23	1.50	-1.65	1.18	2.59
Tāmaki Herenga Manawa	-0.22	1.63	-1.93	0.13	0.07
Tāmaki Herenga Tāngata	0.00	1.94	-2.40	0.52	1.51
Tāmaki Herenga Waka	0.44	1.96	-0.37	0.59	2.47
Taranaki Whanganui Manawatū	0.03	1.55	-1.05	0.81	2.62
Waikato	-0.14	1.66	-1.72	0.62	1.36
Wellington	-0.14	1.65	-2.20	0.49	0.81

Adjusted regional trajectories vary, but they do not follow a single north–south, urban–rural, or large–small region pattern. Table 18 gives total retained assessment record counts across each legacy or refreshed assessment series.

TABLE 18. Total retained assessment record counts by region and assessment series

Region	Mathematics		Reading Comprehension		Reading Vocabulary
	Legacy total 2012 to 2023	Refreshed total 2024 to 2026	Legacy total 2012 to 2024	Refreshed total 2025 to 2026	Total 2012 to 2026
All Aotearoa New Zealand	995,033	307,813	950,679	246,028	559,019
Bay of Plenty Waiariki	67,268	21,134	49,422	13,832	24,888
Canterbury Chatham Islands	103,561	39,019	98,093	27,944	62,281
Hawke's Bay Tairāwhiti	54,278	13,678	56,590	10,821	24,619
Nelson Marlborough West Coast	22,715	9,312	23,770	7,720	17,389
Otago Southland	67,644	21,619	57,045	13,498	36,564
Tai Tokerau	16,926	8,009	15,504	5,752	8,559
Taranaki Whanganui Manawatū	67,984	24,543	58,928	19,683	34,554
Tāmaki Herenga Manawa	139,262	40,294	150,030	36,187	74,375
Tāmaki Herenga Tāngata	134,150	44,515	145,855	43,100	93,702
Tāmaki Herenga Waka	59,740	23,076	69,632	18,971	50,575
Waikato	80,481	23,148	64,917	17,968	45,425
Wellington	181,024	39,466	160,893	30,552	86,088

The amount of data available from each region varies substantially, especially for Reading Vocabulary. Estimates from smaller regional samples should therefore be interpreted with more caution.

The legacy Mathematics series showed mixed change from region to region, though the majority ended the legacy series with similar adjusted average scores to their baseline values from 2012 to 2014. Within the refreshed series, every retained region is above its 2024 reference by 2026, with values ranging from 1.03 in Otago Southland to 2.31 in Bay of Plenty Waiariki. The positive refreshed movement is therefore geographically widespread, although its size varies.

10. Regional patterns

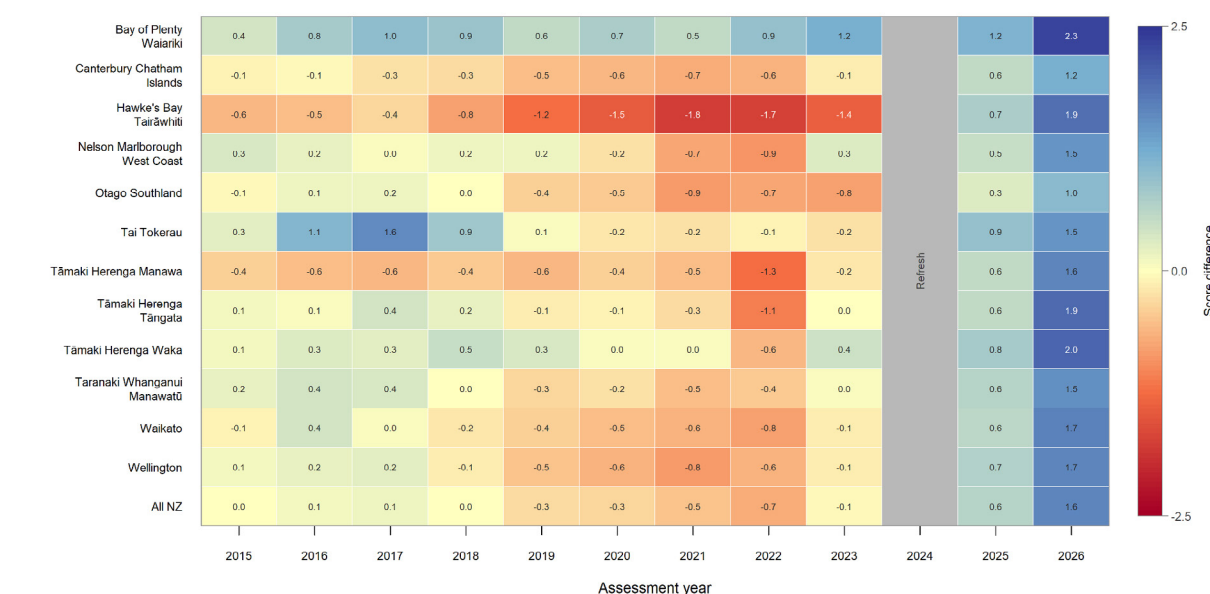


FIGURE 15a. PAT Mathematics regional heatmap

Every retained region is below its own 2012 to 2014 Reading Comprehension reference in 2024. Again, the size differs from region to region, but the negative direction is nationwide. Every retained region then moves positively from 2025 to 2026 within the refreshed series, though some of these movements are not statistically distinct from zero.

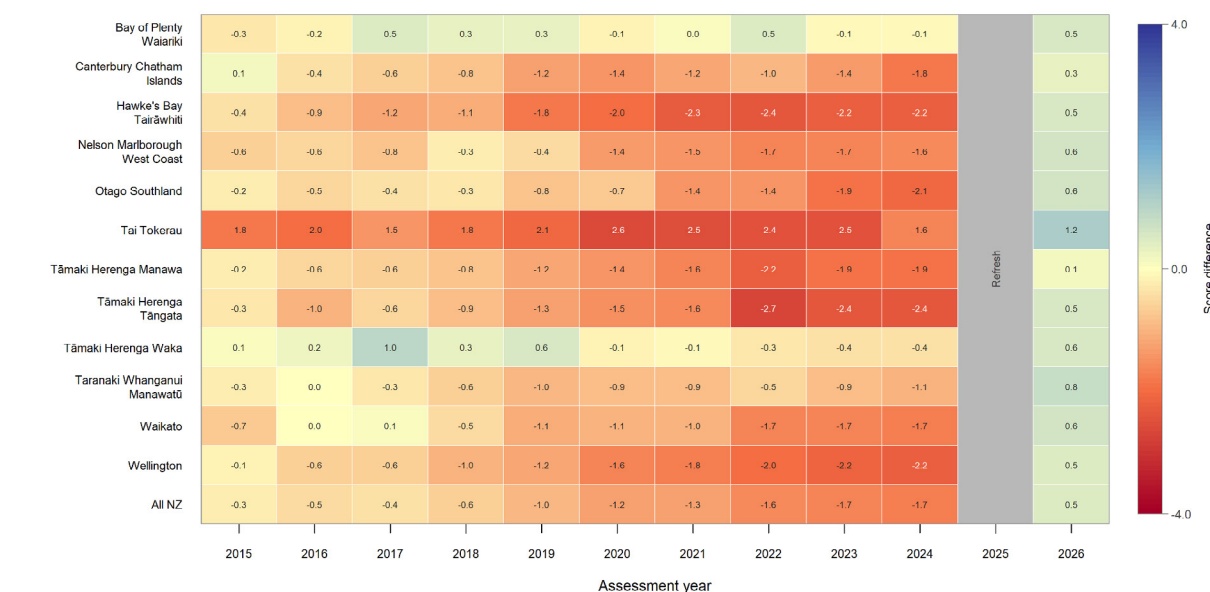


FIGURE 15b. PAT Reading Comprehension regional heatmap

Likewise, every retained region is above its own 2012 to 2014 Vocabulary reference by 2026, although the size ranges from 0.07 in Tāmaki Herenga Manawa to 2.70 in Bay of Plenty Waiariki.

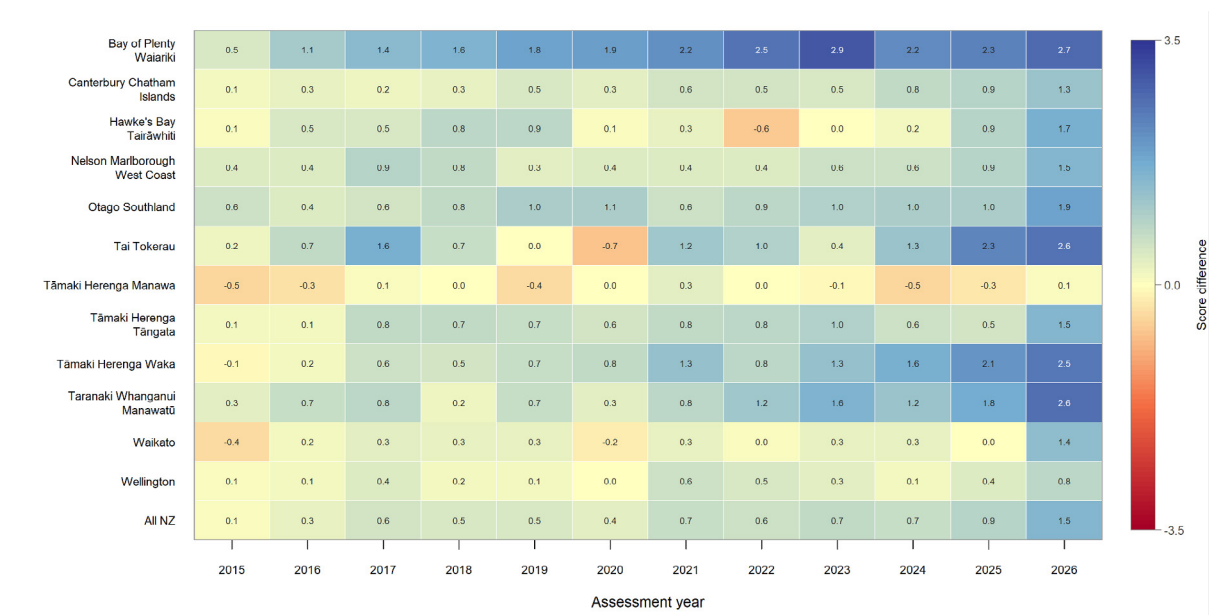


FIGURE 15C. PAT Reading Vocabulary regional heatmap (2012 to 2014 reference)

The region results show that legacy Reading Comprehension decline, early refreshed Mathematics and Reading Comprehension gains, and positive Vocabulary movement were geographically widespread, although their magnitudes varied. This consistency suggests that the national patterns were not driven by changes in only a small number of regions.

The main patterns were geographically widespread. The sensitivity analysis next explores whether their direction and timing also persist under alternative population and modelling choices.

11. Sensitivity and robustness

The main findings in this report come from a model that includes data from all available students, excluding only a small amount of data that represent atypical use of the PAT assessments. This model has many strengths. Students and schools do not need to appear in multiple years, so the model retains mobile students, students tested only once, and schools that use PAT intermittently. Requiring continued participation would produce a more stable sample, but it would also shift the analysis towards a selected group of students and schools that might be less like the full population of PAT users as well as less like the population of all students in Aotearoa New Zealand. The sensitivity analyses presented in this section check whether the findings in this report change when the same PAT data are viewed in more restricted or more flexible ways.

Four sensitivity checks are presented. The “stable students” population includes learners appearing in at least two model-frame years. “Stable schools” require at least five eligible Term 1 records in at least 80% of the years in the relevant segment. “Schools tested every year” restricts the model to schools with eligible records in every year from 2020 to 2026. “School-specific trend” allows schools to follow different trajectories over time rather than modelling them as variations on a single national pattern. Mathematics and Reading Comprehension are checked within their legacy and refreshed segments, while Reading Vocabulary is checked as a continuous series.

These checks are not intended to replace the main model, which was chosen because it has advantages over the alternatives. However, it is useful to assess whether the primary results are similar under other modelling assumptions. Some differences are expected because each check uses a different analytic frame. The important question is whether the same broad direction and timing remain visible under different assumptions. To the extent that this is true, it should build confidence in the results presented elsewhere in the report. These are top-line checks and are not rerun for every subgroup, EQI, entry and transition result, or absolute achievement-level result.

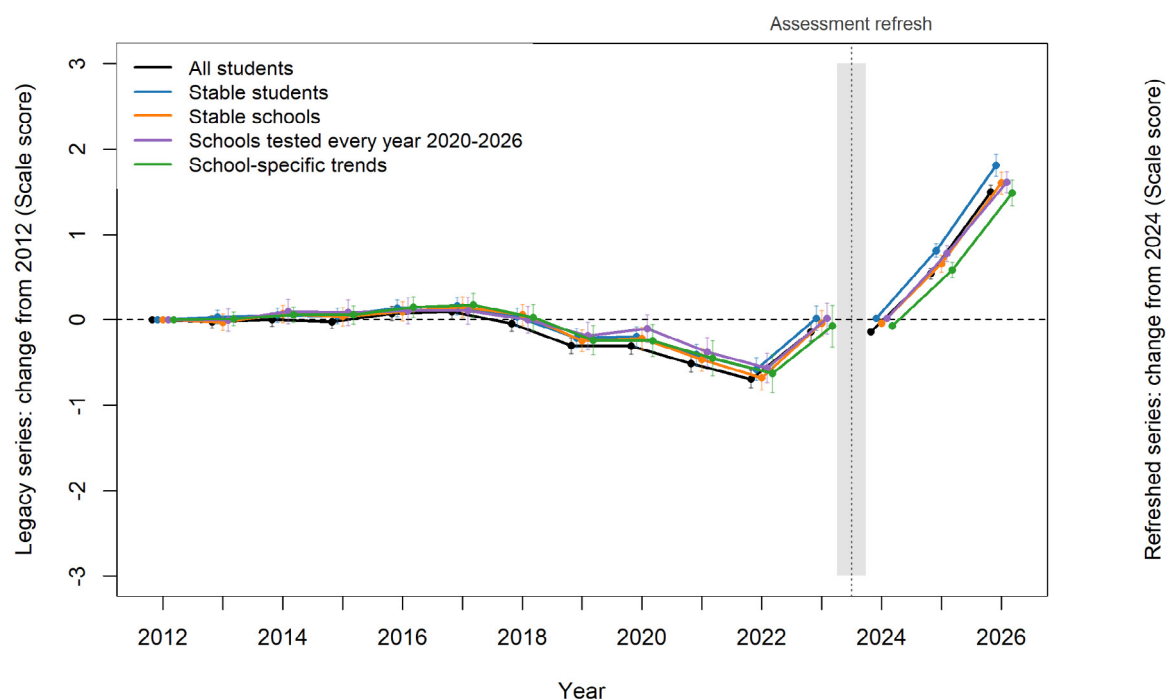


FIGURE 16a. **PAT Mathematics sensitivity comparison**

The Mathematics results are highly stable across the sensitivity models and support the interpretation given elsewhere in this report. The five specifications tell essentially the same story, with the 2022 estimates grouped between -0.69 and -0.56 scale-score points, followed by 2023 estimates that all move back toward the 2012 baseline and range from -0.14 to 0.02. Within the refreshed Mathematics series, the 2026 estimates are also closely grouped, ranging from 1.56 to 1.79 scale-score points above the 2024 refreshed baseline. These are small differences between highly stable models across a range of assumptions rather than competing accounts of the Mathematics trend.

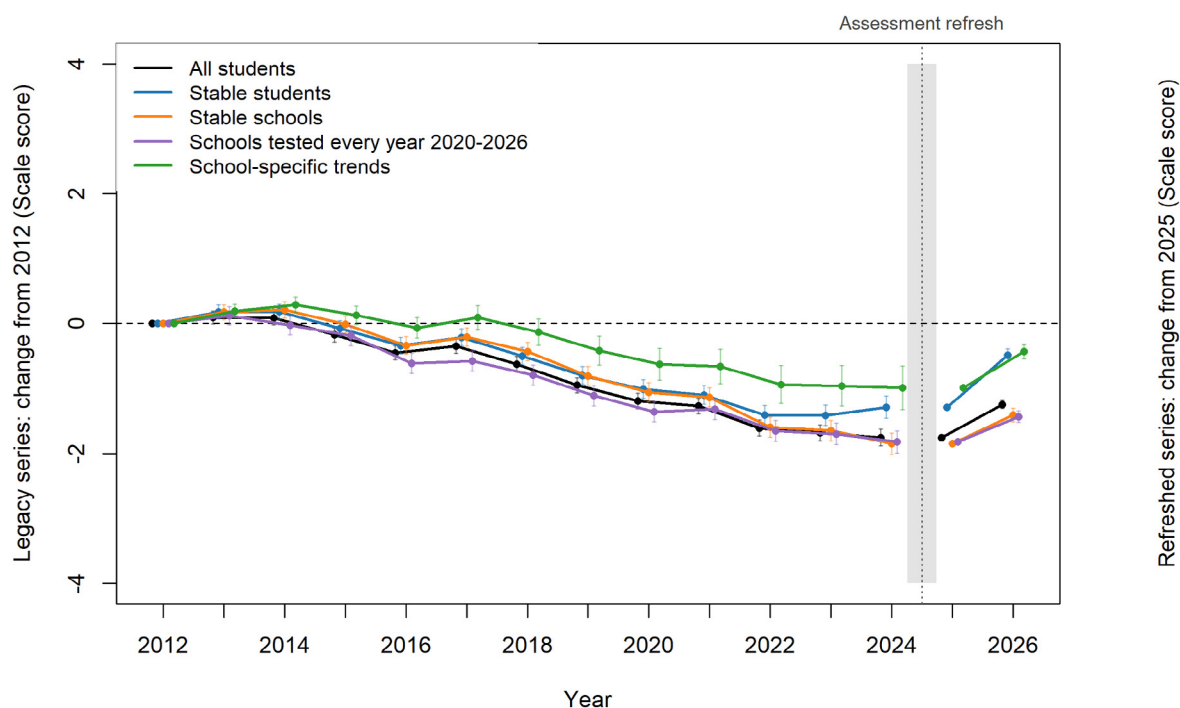


FIGURE 16b. PAT Reading Comprehension sensitivity comparison

While all models agree that there was a broad decline in PAT Reading Comprehension scores, the exact magnitude depends somewhat on the analytic frame. The direction does not change, as every specification shows Reading Comprehension well below baseline at the end of the legacy series. In 2024, the estimates range from -0.99 in the school-specific trends model to -1.85 in the stable schools model. The refreshed series then moves upward in every sensitivity line from 2025 to 2026, with 2026 estimates ranging from 0.39 to 0.80 scale-score points above the 2025 refreshed baseline.

The school-specific trends line is less negative than the other Reading Comprehension sensitivity lines. This suggests that part of the overall decline can be explained by different patterns of school-level change that cannot be captured by the “all students” model but can be modelled at the school level in the “school-specific trends” model. In this model, some of the variation that contributes to the common national trend in the other specifications is instead being absorbed by school-level trend differences. However, even after allowing for those differences between schools, the Reading Comprehension trend remains clearly negative at the student level.

The stable-students estimate is less negative than the all-students estimate, but both show decline. Restricting the frame to this subset of repeatedly tested students changes the magnitude but not the direction.

The sensitivity evidence therefore supports the main Reading Comprehension conclusion. The decline is not dependent on a single model frame. It remains visible when the analysis is restricted to stable students, restricted to stable schools, restricted to schools tested every year from 2020 to 2026, or adjusted to allow school-specific trends. The positive 2026 refreshed movement is also visible across sensitivity checks, but it sits alongside the legacy decline rather than replacing it.

From 2021, direct online records could be separated from manually entered records, an approximate paper proxy. The online share of legacy Reading Comprehension records rose from 51.8% to 70.2% by

2024. Within-school sensitivity estimates suggested that this shift could potentially account for about -0.09 to -0.30 points of the -0.49 scale-score point decline over those years. However, only 33 schools supported a strict like-for-like comparison, and students were not randomly assigned to a mode, so the estimates may also reflect differences between students or testing conditions. Mode is therefore a possible contributor to the later Reading Comprehension decline, but cannot explain the full decline, gender shift, or main ethnicity patterns.

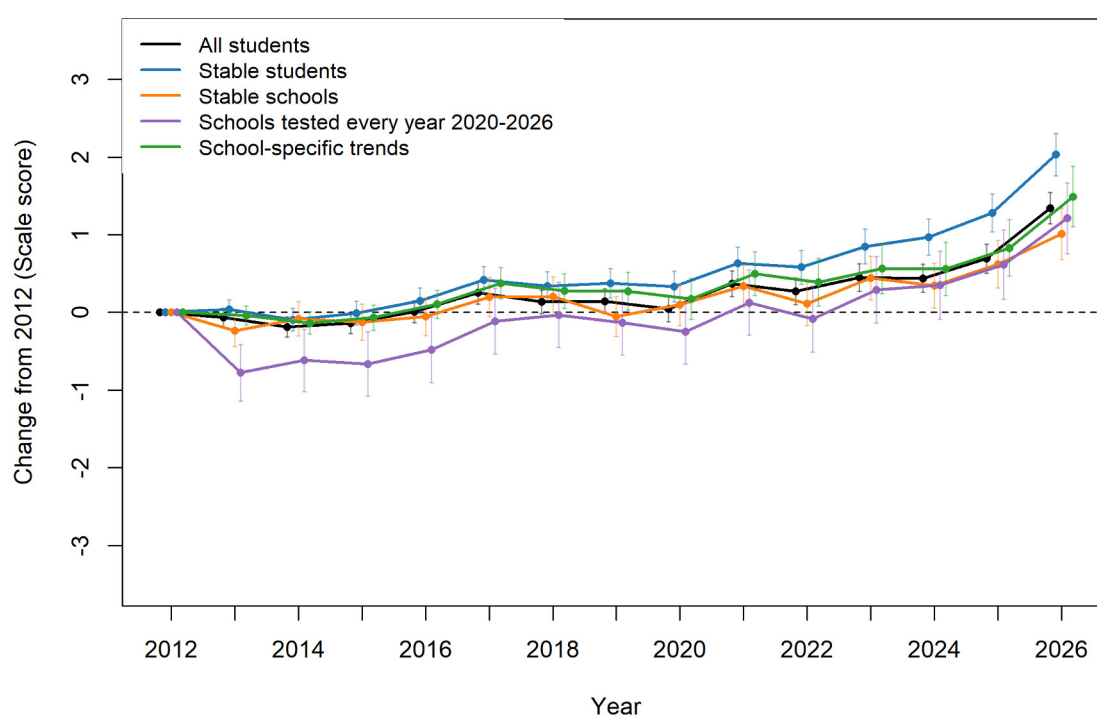


FIGURE 16c. PAT Reading Vocabulary sensitivity comparison

Reading Vocabulary is positive in all five sensitivity lines in the late series. In 2026, the all-students estimate is 1.34 [1.14, 1.54], stable schools is 1.01 [0.68, 1.34], and school-specific trends is 1.49 [1.10, 1.88]. The schools-tested-every-year estimate is 1.21 [0.76, 1.67]. The stable-students estimate is higher, at 2.03 [1.76, 2.30]. Figure 16c shows that the specifications differ in height, but they do not disagree about direction.

In general, the stable-students Vocabulary estimates are higher than those from the other four models, but all five are positive by 2026.

Across the sensitivity analysis frames, Mathematics retains its 2022 low point and 2023 rebound, Reading Comprehension remains negative at the legacy endpoint and positive within the refreshed series, and Vocabulary remains positive late in the series. Those differences that exist between the sensitivity frames are in magnitude rather than direction, and these checks strengthen the principal legacy conclusions and the recent Mathematics direction.

12. Discussion

Fifteen years of PAT results show three different achievement histories. Mathematics weakened late in the legacy series, rebounded, and continued to improve in the refreshed series. Reading Comprehension declined for longer and remains well below the 2012 reference used for the legacy analysis. Reading Vocabulary improved over the period considered in this study.

Most of the main series movements were widespread across many groups, but they were not shared equally across all ages, score rank bands, genders, ethnicity groups, or school contexts. The first part of this discussion draws direct conclusions from the analysis presented in this report. Plausible conclusions that require some speculation beyond the data are considered separately in Section 12.7.

12.1 Three different achievement histories

Mathematics was broadly stable through much of the legacy series before weakening from 2019 and reaching its lowest point in 2022. The substantial 2023 rebound recovered most, but not all, of the decline relative to the earlier legacy reference. Mathematics then rose in both years of the refreshed series and was 1.64 scale-score points, or roughly one-third of a year level, above its 2024 refreshed reference by 2026. The bridging analysis not presented in this report also suggested a small positive movement across the refresh, although that estimate remains less certain.

Reading Comprehension followed a longer negative path. Scores were already below the earlier reference before COVID-19 and fell to their lowest legacy point in 2024. The refreshed series rose by 0.51 points from 2025 to 2026. That early positive movement was not uniform by year level. Year 7 had the largest increase at 0.9 points, while the Year 8 estimate was only 0.1 points and was not clearly different from zero. While the first refreshed estimate was positive, it was not positive enough to suggest that the longer legacy decline has been reversed.

Reading Vocabulary remained one continuous assessment series throughout the period under consideration, moved positively through the later years, and reached its highest point in 2026. The difference from Reading Comprehension is not confined to the national average. Most ethnic groups that declined in legacy Reading Comprehension were nevertheless above their earlier Reading Vocabulary reference by 2026. Among students who took both assessments, the adjusted Vocabulary score relative to the Reading Comprehension score widened by 2.28 points from 2012 to 2024. The two reading measures therefore describe different trajectories at both the population and matched learner levels.

The PAT assessment refreshes prevent a single numerical trend from being drawn across the full period for Mathematics or Reading Comprehension. Legacy and refreshed results answer different within-series questions. Positive movement after a refresh cannot show how much of an earlier decline has been recovered. The longer historical record is nevertheless essential because it shows whether the most recent direction continues, interrupts, or reverses what came before.

Representative national monitoring provides an important check on how far the PAT directions can be generalised. Curriculum Insights found reading in 2025 broadly similar to 2023 at Years 3, 6, and 8. In Mathematics, Year 6 improved from 2024 to 2025 while Years 3 and 8 were broadly stable. Differences

in samples, year levels, assessments, and benchmarks mean these results are not direct replications of PAT. They nevertheless temper a claim of broad system recovery. The refreshed PAT increases are annual signals among PAT users that still require confirmation in representative monitoring (Educational Assessment Research Unit & New Zealand Council for Educational Research, 2026b).

That said, the broad directions of the trends discussed in this report are not products of one model specification. Across five national model frames, Mathematics retains its 2022 low point and 2023 rebound, Reading Comprehension remains negative at its legacy endpoint, and Reading Vocabulary remains positive late in the series. The exact size of the Reading Comprehension decline is more sensitive to modelling choices than its direction. The smaller estimate produced when each school is allowed its own slope indicates that some change treated as a common national decline in the simpler models is instead accounted for by different school trajectories. Those trajectories could reflect teaching and curriculum practice, attendance, student composition, local disruption, or other school circumstances not shared by all schools throughout the country.

12.2 Mathematics improved broadly, with most of the 2026 difference already present at Year 4 entry

Recent Mathematics improvement reached well beyond the national average. Every assessed year level, every ethnic group, and every retained region was above its refreshed reference by 2026. The rank-band analysis also shows improvement across the lower, middle, and upper bands. Regional estimates varied in size, but they did not follow a clear spatial pattern. The positive direction was therefore widespread among the students and schools represented in the PAT model frames.

The age pattern reveals where most of the 2026 difference was already present. The 2026 Year 4 estimate was 2.4 points above its 2024 refreshed reference, compared with 1.1 points at Year 8. The entry and transition analysis reinforces the same conclusion. The Year 4 entry difference was 2.4 points, while the four later transition differences ranged from 0.2 to 0.3 points. All four later intervals excluded zero, but each was much smaller than the entry difference. Most of the recent improvement was therefore already visible when students reached Year 4, with smaller favourable movements added through later year levels.

A national increase led by Year 4 entry directs attention towards experiences before and during the early years of school. However, the smaller positive transition differences still matter. If they recur within the same cohorts, they could help stronger entering groups retain or extend their position. The repeated annual samples cannot yet show that accumulation within one actual cohort.

The distributional result is also broadly positive but not completely even. Legacy Mathematics had no sustained difference between lower, middle, and upper rank-band movement. In the refreshed series, the upper band rose 0.38 points more than the middle by 2026, while the lower Year 8 band rose 0.63 points less than the middle. The improvement therefore reached the score distribution broadly, with some evidence that Year 8 students in the lower rank band shared less fully in the most recent gain.

Widespread improvement did not mean equal improvement across groups. Girls lost more ground as the legacy series weakened and gained less during the refreshed rise. Māori Mathematics moved less favourably relative to its own earlier position, while European and Pākehā students showed negative relative movement in both the legacy and refreshed segments despite the positive refreshed trajectory. These patterns do not overturn the national improvement, but they identify groups for whom the movement was less favourable.

The recent Mathematics pattern was therefore broad, but it was neither uniform nor spread evenly across the school years. That distinction provides a useful contrast with Reading Comprehension, where weakening often appeared early and later movement did not consistently repair it.

12.3 Reading Comprehension decline began early and was not consistently repaired later

Reading Comprehension is the most sustained negative finding. The legacy decline began before COVID-19, reached every assessed year level, and extended across lower, middle, and upper rank bands. Every retained region and almost every ethnic group was below its own earlier reference at the legacy endpoint. The pattern was therefore neither a short pandemic dip nor a decline confined to one age, group, region, or part of the score distribution.

The entry and transition analysis shows that much of the weakening was already present by Year 4. The Year 4 estimate was below its earlier reference from 2015 and was about two points below from 2020 through 2024. Later movement did not simply reproduce that decline. The Year 4 to Year 5 transition became more favourable, moving from 0.21 points below its earlier transition reference in 2015 to 1.22 points above by 2024. This offset part of the weaker entry position but did not remove it.

Across the three PAT subjects, linked learner analyses found only modest catch-up after weaker Year 4 entry. In Reading Comprehension, a cohort one standard deviation below its school's usual Year 4 position was predicted to be about 0.29 scale-score points higher at Year 5 than otherwise expected. The effect was small, did not clearly strengthen over time, and does not show that schools deliberately responded or what information teachers used.

Catch-up was not clearly stronger in schools with a consistent PAT history and was more evident from Year 4 to Year 5 than across later transitions.

The decline also differed by gender. Girls' adjusted Reading Comprehension score fell by 2.4 scale-score points through the legacy series, compared with 1.1 points for boys. This narrowed girls' earlier advantage by 1.3 points. Both groups rose in the refreshed series, but boys rose slightly more.

Partial catch-up between Years 4 and 5 was not consistently maintained. Movement from Year 5 to Year 6 and from Year 6 to Year 7 remained less favourable in 2024, while Year 7 to Year 8 had returned to its earlier reference. The distributional results add another lasting weakness. At Year 8, the upper rank band fell more than the middle from 2017 through 2024. The 2024 upper band was 3.22 points below its reference, compared with 2.41 points for the middle band. The decline therefore began early for many cohorts and remained particularly pronounced among Year 8 students in the upper rank band.

A separate pathway comparison suggests that school transition may be part of the weak Year 6 to Year 7 pattern. Learners moving to intermediate schools made 0.50 to 0.85 points less adjusted progress than learners remaining in full primary schools across all five assessment series, then recovered in legacy Mathematics and Reading Comprehension during Year 7 to Year 8. No clear difference was evident before the transition in the available panels. The Year 7 to Year 8 recovery is consistent with a temporary interruption around the move to intermediate school, but selection remains and the results do not show that intermediate schools caused the initial difference.

The first refreshed comparison is more differentiated than a simple younger versus older account. Every ethnic group had a positive point estimate and every retained region moved positively,

although some regional movements were not statistically distinct from zero. Year 7 had the strongest year-level increase, while Year 8 changed little. The immediate question is therefore whether the refreshed improvement persists and reaches Year 8, not whether it reached all older learners equally in its first year.

Sensitivity results strengthen the conclusion about direction while qualifying its size. The 2024 legacy estimate ranges from -0.99 when trends are allowed to vary by school to -1.85 in the stable schools model. The national direction remains clearly negative after allowing those trends to vary. A geographically widespread decline can therefore coexist with variation in how strongly individual schools changed.

12.4 Reading Vocabulary improved, but lower rank bands and one transition lagged

Reading Vocabulary improvement was widespread. Every retained region and ethnic group was above its own earlier reference by 2026, and every assessed year level ended positively. Years 5 and 6 had the strongest positions relative to their earlier references. All five sensitivity models were positive late in the series. These results support a broad improvement rather than a gain produced by only a few locations or groups.

The age pattern was not a steady rise through each transition. Movement from Year 4 to Year 5 was consistently strong, while Year 6 to Year 7 was repeatedly weak, and movement into Year 8 was usually positive. The most recent data from 2026 show some improvement within this recurring pattern. Year 5 to Year 6 turned positive in 2026, and Year 6 to Year 7 became less negative even though it remained below its earlier reference. The positive endpoint therefore reflects contrasting movements through the primary years rather than equal gains at every step.

The rank-band pattern is a substantial qualification. Lower rank bands changed less favourably than the middle in every year from 2018 through 2026. In 2026, the lower band increased by 0.25 points, compared with 1.24 points in the middle and 1.22 points in the upper band. The lower band still improved, but it shared much less in the gain. Separate Year 4 and Year 8 contrasts were less precise, so the clearest evidence concerns the combined distribution.

The contrast with Reading Comprehension remains one of the most important findings across subjects. Population Vocabulary scores rose while Comprehension scores declined through its legacy series. The same contrast appears across most ethnic groups. Among students who took both assessments, the adjusted Vocabulary score relative to the Comprehension score also widened from 2012 to 2024. The difference is therefore not only a result of different annual assessment populations, although the matched learners remain a selected group. The analysis does not identify the reason.

12.5 Improvement and equity are different questions

A positive national trajectory can be broadly shared without being equally shared. It can also occur while large differences in achievement levels remain. The gender, ethnicity, rank-band, and simulated EQI results all reinforce this distinction. Assessing improvement therefore requires attention both to how groups moved from their own earlier positions and to the adjusted score differences between them.

Across all three assessments, the difference between boys and girls moved in boys' direction, but through three different pathways. In Mathematics, the difference moved 2.42 points in boys' direction

during the legacy series, which is approximately half a year level, and another 0.87 points during the refreshed series. Girls declined more as the legacy series continued and boys gained more during the refreshed rise. Boys' scores were higher at every assessed year level.

In Reading Comprehension, the 1.34 point narrowing between boys and girls occurred because both groups declined and the decline was larger among girls. Girls nevertheless remained higher at every assessed year level. The smaller difference therefore reflected lower scores for girls over time rather than improved scores for both groups. In Reading Vocabulary, the difference shifted by 2.20 points, mainly because boys' scores rose. When year levels were combined, boys' endpoint estimate was slightly higher, although the direction differed by year level. These distinct pathways do not support a single explanation for changing gender differences across the three subjects.

The same movement appeared within major ethnicity groups. Legacy Mathematics and Reading Vocabulary moved towards boys within every retained major group. Legacy Reading Comprehension did so within European/Pākehā, Māori, Pacific, and South, Southeast, and Other Asian groups. The overall gender movement therefore cannot be attributed only to changing ethnic composition.

TIMSS 2023 provides relevant independent context for the Mathematics result. Its Year 5 Aotearoa New Zealand result was the first statistically significant boys' Mathematics advantage in eight cycles (Caygill et al., 2024). The recent Year 5 Mathematics pattern is therefore not unique to PAT. TIMSS 2023 was also the first fully digital cycle, and boys reported greater confidence using digital devices. The independent result supports attention to the PAT pattern without proving that both assessments captured the same underlying change.

Some traditional explanations for gender disparities in education fit the timing poorly. National curriculum policy has required gender inclusive provision since at least 1993, well before the movement observed from 2012 onward (Ministry of Education, 1993). The 2007 curriculum called for education free from sexism and discrimination (Ministry of Education, 2007).

The timing makes explanations based mainly on formal exclusion or openly stated beliefs that girls are less suited to Mathematics less convincing. However, it does not fully rule out more subtle gendered influences. Sixty percent of adults still regarded sexism as a significant issue in Aotearoa New Zealand, while Italian research has associated teachers' implicit stereotypes with larger gender differences in Mathematics and lower confidence among girls (Carlana, 2019; Rangahau Aotearoa Research New Zealand, 2025). Neither source explains the recent Aotearoa New Zealand movement, but they leave open the possibility that contemporary expectations and experiences still matter. These possibilities are discussed further as hypotheses in the speculative Section 12.7.

The ethnicity results require the same distinction between movement and level. Every focal group had a positive point estimate in refreshed Mathematics and refreshed Reading Comprehension, and every group ended above its earlier Reading Vocabulary reference. These widespread positive directions did not produce convergence. At the 2026 endpoint, the adjusted focal group spread was about 11.7 points in Mathematics, which is roughly 2 years of typical PAT progress. North Asian students were about 7.1 points above the next closest focal group. The corresponding Reading Comprehension and Reading Vocabulary spreads were about 4.2 and 4.4 points.

Several profiles show why starting level and movement must be read separately. North Asian students combined a very high adjusted Mathematics level with favourable Mathematics movement, but their first refreshed Reading Comprehension movement was less favourable than for students outside that focal group. European and Pākehā students showed negative relative movement across all three measures even when their refreshed Mathematics and Reading Vocabulary trajectories were positive.

Mathematics scores for Māori students moved less favourably relative to their earlier position. Pacific students declined less than the comparison group in legacy Reading Comprehension, but their first refreshed increase was also smaller. These are different patterns.

School socioeconomic context is associated with differences larger than most annual movements. Across the simulated EQI range, the adjusted score difference is about 10 PAT points, which is greater than 1 year of typical progress in each assessment. This is an association between school context and score, not an individual socioeconomic or causal effect. Holding simulated EQI constant prevents changing school context alone from accounting for the annual trajectories, but it also means that the adjusted ethnicity comparisons do not describe the full inequality created by the uneven distribution of students across school contexts.

The large achievement difference associated with school socioeconomic context was not completely static through time. Across the legacy series, the modelled score gap between lower and higher barrier school contexts narrowed by 0.65 points in Mathematics, 1.18 in Reading Comprehension, and 1.58 in Reading Vocabulary. These changes were modest relative to the size of the remaining gap, but each was statistically significant. Importantly, this narrowing arose through different patterns of achievement change.

In legacy Mathematics, scores declined in the lower barrier context while remaining broadly stable in the higher barrier context. Reading Comprehension declined in both contexts but declined more in the lower barrier context. These gaps narrowed because outcomes became less favourable where scores had previously been higher, not because outcomes improved in the higher barrier context. Reading Vocabulary followed a more positive pattern. Scores were broadly stable in the lower barrier context and improved in the higher barrier context. A narrowing school context gap therefore cannot be interpreted on its own as improved equity. What matters is whether it reflects gains among students facing greater barriers or declining achievement among those facing fewer.

12.6 What the historical timing and evidence can establish

The Mathematics timeline places a clear limit on the policy interpretation. The rebound was already visible in the legacy series in 2023, before the daily hour requirement began in 2024 and the revised Mathematics content became compulsory in 2025. These settings therefore cannot explain the beginning or the full pattern of improvement, although further positive movement continued after their implementation.

The later observations are nevertheless relevant to assessing whether the new settings helped sustain that movement. Term 1 2025 is the first PAT point that could reflect a completed school year under the daily hour requirement. Term 1 2026 is the first after the revised Mathematics content had been required in Years 0 to 8 for a full school year (Ministry of Education, 2024c). By 2025, ERO reported broad uptake of the revised content, accompanied by changes in teaching, new resources, and professional learning (Education Review Office, 2025). This evidence establishes that implementation was under way during the later period, but not whether it contributed to the PAT results or by how much.

The Reading results are less able to separate movement that predated the new policy settings from movement that followed them. The refreshed Reading Comprehension series begins in 2025, so it has no comparable observation from before the daily hour requirement. Reading Vocabulary spans both periods but was not a separate policy target, meaning that any effect would have operated indirectly through reading or wider language instruction. The three subject trajectories therefore cannot support a single conclusion about the effects of the foundational learning changes.

The 2026 policy context could potentially have affected the PAT results through participation changes as well as teaching. From that year, schools were required to assess students in Years 3 to 8 twice annually using one of several specified tools, including PAT (Ministry of Education, 2025a). PAT itself was not compulsory, but the requirement may have changed which schools used PAT and which students they assessed. Positive movement remained in all three measures when the comparison was restricted to school and year-level combinations represented in both 2025 and 2026. Changes in school participation and year-level coverage therefore do not appear to be the main explanation. Students' prior PAT histories also changed, but these are too closely tied to year level, selection, and school testing practice to identify a repeat testing effect. The 2026 results should therefore be interpreted with some additional caution, although the available checks do not show that changing participation produced them.

Administration shifted gradually from paper to online. NZCER's mode equivalence study found broadly similar Reading Comprehension item behaviour across modes but lower scores online on average, with average differences of roughly 2 to 5 scale points and slightly larger effects for Māori students and students in decile 1 and 2 schools (Eyre et al., 2017). Exploratory modelling suggests that mode could account for part of the negative Reading Comprehension movement from 2021 to 2024. However, these are observational sensitivity calculations and, in any case, cannot explain the full decline from 2012, the movement in boys' direction, or the main ethnicity trajectories. Refreshed estimates changed little after mode adjustment. Mode is therefore a plausible partial influence on the later legacy Reading Comprehension trajectory, not a demonstrated or complete explanation.

Policy timing, assessment participation, and testing mode therefore help interpret particular parts of the results, but they do not provide one explanation across all three subjects. The next section considers other possible explanations for the remaining patterns.

12.7 Plausible explanations and hypotheses

The explanations in this subsection are hypotheses for further testing. Each is connected to an observed pattern, but none can be established by the PAT analyses alone. The evidence that fits each explanation is considered alongside evidence that limits it.

12.7.1 Post COVID-19 cohort replacement and recovery

One plausible interpretation of the recent Mathematics result concerns the experience of the 2026 Year 4 cohort. A typical student in that cohort was in Year 1 in 2023, later in the recovery period than the cohorts that reached Year 4 in the preceding years. Those earlier cohorts began school during 2020, 2021, or 2022 and encountered systemic issues with closures, illness, isolation, and low attendance. The 2026 cohort was not untouched by the pandemic because much of its early childhood education still occurred during that period, and disruption did not end when schools reopened. Its formal school years nevertheless began without the main COVID-19 school closures or the immediate aftermath.

Schools were still working in a recovery environment when this cohort entered Year 1. The Education Review Office (ERO) reported catch-up lessons and described practices that schools could use to locate gaps, adapt teaching, and accelerate learning. The report also described strengthened work with whānau and monitoring engagement to identify students at risk of disengaging (Education Review Office, 2021).

By 2023, this local activity sat within a wider concern about recovery. ERO found that only 19% of principals believed their schools had recovered from the COVID-19 disruption, while 35% of surveyed

learners said they needed to catch up in Mathematics. ERO called for schools to identify learning gaps, improve attendance, and accelerate progress in reading, writing, and numeracy (Education Review Office, 2023). If many schools responded to the same problem, varied local efforts could together have influenced results across the system.

Early literacy provides an example of a recovery response with substantial reach. By Term 1 of 2023, Better Start Literacy Approach training had reached or was beginning in 810 schools. Its COVID-19 response extension to Year 2 was expressly intended to support children affected by disruption (Novak, 2023). The evidence is less direct for Mathematics. The main formal lost learning initiative focused largely on Years 7 to 13, so it would be too strong to claim that this initiative caused the stronger Year 4 result (Education Review Office, 2023).

Attendance may have contributed to the Mathematics pattern. Regular attendance began declining around 2015, fell particularly sharply in 2022, and then recovered through 2026 (Ministry of Education, 2026a). That timing fits the Mathematics low point and rebound. Ministry analysis finds a stronger relationship between attendance and Mathematics than between attendance and reading (Webber, 2020). Attendance moved similarly for boys and girls during the recovery, so it does not explain the gender shift. It may also have contributed to Reading Comprehension decline, but it does not explain why that decline began earlier, lasted longer, and differed from Reading Vocabulary.

Early oral language provides a narrower possibility for recent reading entry cohorts. ERO reported that 20% of children were struggling with oral language at age 5. Fifty-nine percent of surveyed early childhood teachers and 65% of new entrant teachers considered COVID-19 to have affected children's language development (Education Review Office, 2024). This could have contributed to weak Year 4 entry among more recent cohorts and to unequal effects across school contexts. It cannot explain a Reading Comprehension decline visible from 2015. Boys were also more likely to experience early oral language difficulty, so this evidence does not explain girls' greater later decline.

12.7.2 Possible contribution from later policy exposure

The same cohort encountered several Mathematics-focused changes relatively early in schooling. It experienced the daily hour requirement during Years 2 and 3, then completed Year 3 under the revised Mathematics content with new classroom resources and widespread professional learning. The curriculum was published in October 2024 for required use in 2025, and ERO later documented broad uptake and related changes in teaching (Education Review Office, 2025; Ministry of Education, 2024b). The Term 1 2026 PAT result therefore followed a full school year in which these settings could have influenced learning.

This timing is compatible with a later policy contribution, but it cannot separate that contribution from recovery, attendance, or local support efforts after COVID-19. It also cannot explain the 2023 rebound, which predates the education reforms. An account based on policy alone is therefore inconsistent with the full history, while an account in which later implementation helped sustain existing improvement remains possible.

The next Year 4 cohorts provide a useful test. The cohort reaching Year 4 in 2027 began school in 2024 and experienced the daily hour requirement from Year 1. The 2028 cohort began school in 2025 and experienced both the daily hour and revised Mathematics teaching from entry. Continued strength across these cohorts would fit a more durable improvement or cumulative policy exposure better than an effect limited to one cohort. A single 2026 jump followed by flatter entry results would fit temporary cohort replacement more closely.

If improvement is being sustained through schooling, it may become more visible through repeated positive transition differences than through another large Year 4 jump. The refreshed Mathematics transition estimates were about 0.2 points at each later step. They are modest in one interval but could become educationally important if they recur across years within the same cohorts. The current analysis cannot yet establish that recurrence.

12.7.3 School responses to weaker Year 4 entry

The increasingly favourable Year 4 to Year 5 Reading Comprehension transition alongside the worsening initial position of Year 4 cohorts as they enter the testing window is compatible with schools responding to weaker entering cohorts, but the data cannot establish whether they actually did so or what information prompted the response.

The consistent PAT comparison addresses only whether conditional convergence was stronger in schools with a consistent PAT history, and there was no clear evidence that it was. This remains a weak test of the broader possibility that schools identified and responded to weaker cohorts. Schools could have drawn on e-asTTLe, classroom assessment, teacher observation, or other monitoring not captured here. The result shows only that consistent PAT history was not associated with stronger convergence. It does not show that no school responded to identified need.

12.7.4 Gender shifts

The movement in boys' direction across Mathematics, Reading Comprehension, and Reading Vocabulary calls for explanations that account for both its breadth and its different forms. Some traditional paradigms, such as accounts centred on girls' exclusion from Mathematics or stereotypes that girls are less capable in Mathematics, cannot explain the two reading results. They also fit poorly with changes developing long after gender-inclusive curriculum expectations were established.

Independent evidence provides context for both Mathematics and reading. TIMSS 2023 found Aotearoa New Zealand's first statistically significant Year 5 boys' Mathematics advantage in eight cycles (Caygill et al., 2024). PIRLS 2021 found that the Year 5 reading difference remained in girls' favour but was the smallest in 20 years. The longer term narrowing largely reflected a slight weakening among girls (Chamberlain & Forkert, 2023). The PIRLS result parallels the direction of the PAT Reading Comprehension movement. Neither comparison independently tests the Reading Vocabulary trajectory. These parallels show that the Mathematics and Reading Comprehension directions are not isolated to PAT, but they do not establish a cause.

Different, contemporary gendered experiences remain possible influences, but they need not operate in the same way in each subject. For example, Carlana's Italian study links teachers' implicit stereotypes with wider Mathematics differences and lower confidence among girls, but it does not establish this cause in Aotearoa New Zealand or in reading (Carlana, 2019). Reading research links comprehension with reading frequency, attitudes, and competence beliefs, which often differ between girls and boys (Logan & Johnston, 2009). Vocabulary and reading comprehension share some language foundations but should be treated as related rather than interchangeable constructs (Mol & Bus, 2011; NZCER Assessment Services, 2024). These studies identify possible pathways rather than causes of the observed change.

Aotearoa New Zealand evidence also shows that explicit expectations and confidence need not point in the same direction. In a recent study in three Aotearoa New Zealand intermediate schools, teachers expected girls to achieve more highly in Mathematics, while boys reported higher Mathematics self-efficacy at the end of the year. The study drew on three urban schools rather than a national sample

and found no statistically significant gender difference in mathematics achievement (Watson & Rubie-Davies, 2024).

The divergent gender score trajectories across assessment subjects make a single, simple explanation unlikely to be responsible. Mathematics shifted because girls declined more and boys later gained more. Reading Comprehension shifted because both groups declined but girls declined more. Reading Vocabulary shifted through boys' gains, but there was no corresponding decline among girls. A broad change in engagement, attendance, access to learning, assessment behaviour, digital activity, or experiences in and outside school could have affected all three assessments while interacting with subject-specific teaching and learning in different ways. Separate subject-specific processes could also have produced movement in the same direction.

Distinguishing these possibilities requires repeated achievement measures linked to evidence about how students experience and engage with learning in and beyond the classroom. Future studies should explore whether these experiences changed differently for girls and boys and whether they help explain subject-specific movement or a process shared across subjects.

12.7.5 Reading Vocabulary and Reading Comprehension divergence

The literacy story told by the past 15 years of PAT assessments is one of divergence between Reading Comprehension and Reading Vocabulary. The difference also widened among learners who completed both assessments in the same calendar year and at the same year level. This makes it hard to explain the overall divergence only by changes in who was assessed.

One possibility is that students encountered more words while receiving less sustained practice in constructing meaning across connected written text. The expansion of phones and other connected devices could have contributed by increasing exposure to short, varied, and often audiovisual content. The point at which students gain independent access through their own phone may be particularly important because it makes digital content more portable and readily available. Aotearoa New Zealand does not have a strong national time series showing when children first acquired smartphones. In 2018, however, 94% of children aged 9 to 11 used the internet at home at least weekly and almost half used a smartphone to go online at least weekly. This is unlikely to have diminished since then. School survey evidence from before the pandemic already placed phone ownership at about one-quarter in Year 4 and about two-thirds in Year 7 (Pacheco & Melhuish, 2019).

There is some evidence that digital content can support word learning without supporting comprehension to the same extent. A meta-analysis of digital storybooks found that features focused on word meanings could improve vocabulary while providing no benefit to, and sometimes interrupting, understanding of the wider story (Furenes et al., 2021). This shows how word learning and comprehension can separate under some digital conditions. It does not show that ordinary phone use strengthened PAT Reading Vocabulary. PISA 2018 recorded declining reading enjoyment, less frequent reading across several print forms, and increasing online reading. PIRLS 2021 found that Year 5 reading attitudes and confidence had weakened since 2016. It also showed that almost all students had access to digital devices during reading lessons (Chamberlain & Forkert, 2023). Wider evidence links sustained print exposure with both vocabulary and comprehension and finds that leisure digital reading does not consistently support comprehension in the same way (Altamura et al., 2025; Mol & Bus, 2011).

The evidence therefore does not support a simple claim that more digital device exposure produces stronger vocabulary. Indeed, a recent Aotearoa New Zealand longitudinal study associated greater early childhood screen exposure with slightly lower vocabulary and early educational skills after

accounting for family background (Gath et al., 2026). However, it is possible that device exposure changes the form and pattern of later language acquisition, partially decoupling the learning experience of encountering words from acquiring sustained reading comprehension.

12.7.6 Generative AI and wider digital change

Generative AI could have contributed to the recent rise in PAT Mathematics by changing the explanations and resources available to teachers and students. In late 2024, many teachers in a voluntary study focused on Years 5 to 8 were already reporting that they used AI for planning and adapting learning materials, although respondents had a strong prior interest in AI (Coblentz et al., 2025). By Term 1 2026, more than four in five teachers surveyed by ERO reported using AI, mainly for similar purposes (Education Review Office, 2026). These practices could have affected younger students through changes in teaching, even where students did not use AI themselves.

12.7.7 Migration patterns

Changing migration history may help explain parts of some ethnicity trajectories. The PAT ethnicity categories include students whose families have very different settlement histories. Some arrived recently, while many others were born in Aotearoa New Zealand to parents who arrived not long before them. Changes within an ethnicity category may therefore reflect changes in family migration history as well as achievement. In PISA 2022, 29% of Aotearoa New Zealand students had an immigrant background, but only 14% were themselves born overseas. Students with an immigrant background scored more highly in Mathematics on average, while their Reading difference was not clear (OECD, 2023). PIRLS 2021 also found considerable variation in the backgrounds of Asian Year 5 students, around two-thirds of whom were born in Aotearoa New Zealand (Chamberlain, 2024).

The timing offers some support for the possibility that migration patterns may help explain some ethnicity trajectories. High migration from 2014 to 2019 could have affected the PAT population immediately through children arriving with their families and later as Aotearoa New Zealand-born children of recent migrants reached the assessed year levels. The exceptional 2023 inflow would mostly affect the 2024 and later results because PAT testing occurred in Term 1. Census growth was especially rapid in parts of the Latin American and Asian populations over this period (Stats NZ, 2024). This broadly overlaps with some of the more positive Mathematics and Vocabulary movement among smaller migrant-associated groups.

The correspondence is not close enough to support a stronger conclusion. There was no shared change among migrant-associated groups when the border closed or reopened. Some trajectories also began well before those events, while the refreshed Mathematics increase extended across every ethnic group. Changing migration composition is therefore a plausible contributor to some ethnicity results, particularly for smaller and rapidly changing groups, but it does not explain the main national patterns.

12.8 What should be explored next

The findings point to several questions that may deserve more attention. Some will become clearer over the coming years as the refreshed PAT series continue to gather more data. Others require evidence that PAT cannot provide on its own, particularly about what changed in classrooms and which learners experienced those changes.

It is important to determine whether the recent Mathematics improvement persists. The 2026 increase was strongest at Year 4, although smaller positive differences were also visible through the later transitions. The cohort assessed in Year 4 in 2026 should therefore be followed as it reaches Year 5 in 2027 and later year levels. Its progress should be considered alongside the entry results for the next Year 4 cohorts. Continued strength at entry and through later transitions would support a durable improvement. A large result confined to one or two entering cohorts would fit a more temporary recovery or cohort replacement explanation.

The 2028 Year 4 cohort will be particularly informative. These students began school when the revised Mathematics content and structured literacy changes were introduced in 2025. Their results will not provide a simple test of policy effectiveness because several changes occurred at the same time. They will, however, provide the first broad PAT entry evidence for a cohort exposed to the newer settings from the beginning of school.

For Reading Comprehension, an immediate question is whether the positive movement from 2025 to 2026 continues. Year 8 changed little in the first refreshed comparison, while Year 7 showed the strongest increase. It will be important to see whether that stronger Year 7 result appears at Year 8 one year later. Later Year 4 cohorts will also show whether the long decline in entry scores begins to reverse. Several refreshed years will be needed before a sustained change can be convincingly distinguished from more ordinary annual variation.

Interpretation of recent policy should move beyond implementation dates. It would be ideal to pair PAT or SMART evidence with more detailed school-level information about what schools actually changed and when those changes reached students. Measures of curriculum uptake and teaching time could be linked with later achievement while taking account of schools' earlier trajectories. An important question is not simply whether scores rose after a policy began. It is whether schools with earlier or stronger implementation changed differently from otherwise similar schools. Even then, the evidence would remain observational and would need careful interpretation.

The school-specific trend results presented in this report show that schools did not all follow the national Reading Comprehension trajectory to the same degree. Schools with persistently stronger or weaker trajectories could be studied more closely to identify the conditions associated with those differences. This work should use several years of evidence rather than selecting schools because of one unusual result. It could also help connect the national patterns with more detailed evidence about teaching and attendance.

The divergence between Reading Vocabulary and Reading Comprehension remains one of the clearest unresolved findings. Further work should examine whether the separation is more concentrated in particular kinds of text or at particular ages. Linked measures of reading fluency and sustained comprehension could help show whether students are learning more words without a corresponding improvement in understanding connected text. Oral language and background knowledge may also help explain why the two measures have moved differently. Item-level analysis could also test whether the divergence is associated with particular comprehension demands.

The movement in boys' direction also would benefit from focused follow-up investigation. One task is to assess whether changing item behaviour or response patterns contributed to the result presented here. Another, broader question is whether girls and boys are experiencing differing changes in confidence or engagement that can credibly be shown to have likely emerged in recent years. The best explanations will involve both explanation of the current difference as well as a plausible historical account of how it slowly came into being. The three PAT subjects may need to be

considered separately as the explanations quite likely differ among them, either in kind or in degree. Mathematics and Reading Comprehension changed mainly through less favourable movement among girls. Reading Vocabulary changed mainly through stronger gains among boys. A single explanation is unlikely to account adequately for all three patterns.

The weaker Year 6 to Year 7 progress among students moving to intermediate schools should also be explored with stronger experimental designs or more comprehensive linked data beyond the scope of this study. Linked learner data could follow students before and after the move and compare them with similar students who remain in full primary schools. Attendance information around the transition would be especially useful. The present results are consistent with a temporary interruption, followed by some recovery in the next year, but they cannot show whether the intermediate schooling transition itself contributed to the weaker progress in some way.

Equity monitoring should continue to distinguish improvement from convergence. Groups may improve from their own earlier position while large differences in achievement remain. Individual measures of socioeconomic circumstances would provide information that school-level EQI cannot capture. The lower part of the Reading Vocabulary and Year 8 Mathematics distribution warrant more attention because they shared less fully in recent gains. Future evidence could establish whether these learners begin to participate more fully if the positive national directions continue.

Smaller ethnicity groups would benefit from more precise information about migration and language background. Broad ethnicity categories may sometimes contain students with very different histories of settlement and schooling. Changes in those histories could influence some of the more distinctive Mathematics and Vocabulary trajectories. Larger samples and linked contextual information would help better distinguish changing population composition from changing achievement among otherwise similar students.

Changes in assessment practice should remain part of future monitoring. Participation and prior PAT exposure may change if the new twice-yearly assessment requirement becomes established.

Having more than one credible source of student achievement data is a strength of Aotearoa New Zealand's monitoring environment. PAT provides high-volume annual evidence from ordinary school use. Curriculum Insights provides representative national monitoring, while international studies add broader points of comparison. When these sources point in the same direction, confidence increases that the movement reflects wider system change rather than features of one assessment. When they differ, the disagreement is also informative. It may reveal differences in who is assessed, what is measured, or when change became visible. Taken together, complementary sources provide a stronger account of achievement change than any one source could provide alone.

No single study can answer all of these questions. Longer assessment series will show whether the most recent movements persist and how they travel through successive cohorts. More detailed research may clarify what changed in learning conditions and whether recent gains were shared by students who had benefited less. The strongest account of achievement change will emerge from the way these findings align, or fail to line up, with other credible sources of evidence, rather than from any one measure or explanation.

13. Conclusion

This report adds a 15-year annual perspective to the evidence on student achievement in Aotearoa New Zealand. Its central finding is that Mathematics, Reading Comprehension, and Reading Vocabulary followed three different histories. The PAT evidence therefore does not support a single account of general decline or recovery.

PAT Mathematics was broadly stable through much of the 2010s, weakened from 2019, and reached its lowest legacy point in 2022 after COVID-19. Most of that loss had been recovered by 2023. The refreshed series then rose in 2025 and 2026. This recent movement was widespread across year levels, regions, and ethnic groups, although it was not evenly shared. The largest 2025 and 2026 differences were already present when students entered the PAT window in Year 4. Smaller positive differences were also visible through later transitions. The durability of the entering cohort results, and its movement through later year levels, will be central to understanding whether Mathematics has entered a sustained period of improvement.

Within the three PAT subject assessments considered here, Reading Comprehension shows the largest and most sustained decline. Its decline began before the pandemic, continued through the final legacy year, and was visible across the country and throughout the score distribution. Much of the decline that has developed over time is already present by the time students have first been tested in Year 4. More favourable movement from Year 4 to Year 5 offset some of that weaker entry position but did not reverse it. Progress also remained weaker through parts of the middle primary years. The positive movement from 2025 to 2026 within the refreshed series is important, but one interval is not enough to establish a persistent change in direction.

Reading Vocabulary followed a more positive course. It ended above its earlier reference at every assessed year level and reached its highest national point in 2026. That improvement was not uniform. Students in the lower rank band shared less fully in the gain, and progress from Year 6 to Year 7 remained comparatively weak. Most importantly, Vocabulary and Reading Comprehension moved apart. The same divergence appeared among students who took both assessments. Broad claims about reading achievement can therefore conceal quite different changes in the language knowledge and comprehension captured by the two assessments.

Taken together, the PAT record does not support a picture of achievement moving uniformly towards decline or recovery. It shows three subjects moving apart. Reading Comprehension declined substantially during the legacy series and remains a concern despite the positive first refreshed movement. Mathematics has rebounded broadly, but that improvement has not yet proved durable. Reading Vocabulary has risen, though not equally for all learners. The next few years will show whether the most recent gains foreshadow lasting increases in assessment scores.

References

- Altamura, L., Vargas, C., & Salmerón, L. (2025). Do new forms of reading pay off? A meta-analysis on the relationship between leisure digital reading habits and text comprehension. *Review of Educational Research*, 95(1), 53–88. <https://doi.org/10.3102/00346543231216463>
- Bethhäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7, 375–385. <https://doi.org/10.1038/s41562-022-01506-4>
- Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3), 1163–1224. <https://doi.org/10.1093/qje/qjz008>
- Caygill, R., & Borthwick, R. (2024). *TIMSS 2023 mathematics Year 9: Trends over nearly three decades of TIMSS findings*. Ministry of Education.
- Caygill, R., Borthwick, R., & Bennett, H. (2024). *TIMSS 2023 mathematics year 5: Trends over nearly three decades of TIMSS findings*. Ministry of Education.
- Caygill, R., & Zhao, B. (2022, January). *Progress and achievement and the context of mathematics and statistics learning in New Zealand (English medium education)*. Ministry of Education.
- Caygill, R., Zhao, B., Hunter, H., & Park, S. (2022, January). *How our education system is performing for literacy*. Ministry of Education.
- Chamberlain, M. (2024, October). *PIRLS 2021: Asian students in year 5—a diverse group*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/schooling/asian-students-in-year-5a-diverse-group>
- Chamberlain, M., & Forkert, J. (2023, May). *Reading literacy at year 5: New Zealand's participation in PIRLS 2021*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/2539/pirls-progress-in-international-reading-literacy-study-202021/reading-literacy-at-year-5-new-zealand-participation-in-pirls-2021>
- Coblentz, D. (2026). *COVID-19 and student achievement in Aotearoa New Zealand—Progressive Achievement Test evidence on mathematics and reading comprehension*. New Zealand Council for Educational Research. <https://doi.org/10.18296/rep.0095>
- Coblentz, D., Dong, J., & Gibbs, B. (2025). *Generative artificial intelligence in Aotearoa New Zealand primary schools: Teacher and student survey findings*. New Zealand Council for Educational Research. <https://doi.org/10.18296/rep.0077>
- Darr, C. (2024, July 18). *PAT Pāngarau | PAT Mathematics: Changes to scale scores and national reference information*. New Zealand Council for Educational Research. <https://www.nzcer.org.nz/news-and-blogs/pat-pangarau-pat-mathematics-changes-scale-scores-and-national-reference-information>
- Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, 25, 23–38. <https://doi.org/10.1016/j.edurev.2018.09.003>
- Education Gazette. (2023, August 14). *Model underpins literacy & communication and maths*. <https://gazette.education.govt.nz/articles/model-underpins-literacy-and-communication-and-maths/>

- Education Gazette. (2024, February 22). *Supporting change in our schools and kura*. <https://gazette.education.govt.nz/articles/supporting-change-in-our-schools-and-kura/>
- Education Review Office. (2021, December 6). *The impact of COVID-19 on teachers and principals*. <https://www.evidence.ero.govt.nz/documents/the-impact-of-covid-19-on-teachers-and-principals-december-2021>
- Education Review Office. (2023, June 22). *Long COVID: Ongoing impacts of COVID-19 on schools and learning*. <https://www.evidence.ero.govt.nz/documents/long-covid-ongoing-impacts-of-covid-19-on-schools-and-learning>
- Education Review Office. (2024, August 22). *Let's keep talking: Oral language development in the early years*. <https://www.evidence.ero.govt.nz/documents/lets-keep-talking-oral-language-development-in-the-early-years>
- Education Review Office. (2025, October 15). *A new chapter: How well are the changes to English and maths going?* <https://www.evidence.ero.govt.nz/documents/a-new-chapter-how-well-are-the-changes-to-english-and-maths-going-research-report>
- Education Review Office. (2026, July 30). *Ready or not: How are schools responding to artificial intelligence? Summary report*. <https://www.evidence.ero.govt.nz/documents/ready-or-not-how-are-schools-responding-to-artificial-intelligence-summary-report>
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2020). *National Monitoring Study of Student Achievement: English 2019—key findings* (NMSSA Report 22). Ministry of Education. https://nmssa-production.s3.amazonaws.com/documents/2019_NMSSA_ENGLISH.pdf
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2023). *Mathematics and statistics 2022: Achievement findings* (NMSSA Report 30). https://nmssa-production.s3.amazonaws.com/documents/NMSSA_2022_Mathematics_Achievement_Report.pdf
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2024a, October). *Curriculum Insights and Progress Study: Foundation area—Mathematics and Reading results 2023*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/schooling2/curriculum/curriculum-insights-and-progress-study/foundation/curriculum-insights-and-progress-study-foundational-area-mathematics-and-reading-results-2023>
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2024b, August). *Curriculum Insights and Progress Study: Foundation area—Reading 2023*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/schooling/curriculum-insights-and-progress-study/foundation/curriculum-insights-and-progress-study-foundational-area-reading-2023>
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2026a). *Curriculum Insights and Progress Study: Contextual insights using statistical methods report 2023, foundation areas: Mathematics and reading*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/curriculum-insights-and-progress-study/foundation/contextual-insights-using-statistical-methods-report>
- Educational Assessment Research Unit, & New Zealand Council for Educational Research. (2026b). *Curriculum Insights and Progress Study: Summary of the 2025 foundation assessments: Mathematics, reading and writing*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/curriculum-insights-and-progress-study/foundation/curriculum-insights-and-progress-study-summary-of-the-2025-foundation-assessments>

- Educational Measurement and Assessment. (2025, August). *A short synthesis of findings from our large education studies*. Ministry of Education. https://www.educationcounts.govt.nz/__data/assets/pdf_file/0015/221073/A-short-synthesis-of-findings-from-our-large-education-studies.pdf
- Egert, F., Cordes, A.-K., & Hartig, F. (2022). Can e-books foster child language? Meta-analysis on the effectiveness of e-book interventions in early childhood education and care. *Educational Research Review*, 37, Article 100472. <https://doi.org/10.1016/j.edurev.2022.100472>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127. <https://doi.org/10.1037/a0018053>
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30(2), 326–343. <https://doi.org/10.1111/j.2044-835X.2011.02042.x>
- Eyre, J., Berg, M., Mazengarb, J., & Lawes, E. (2017). *Mode equivalency in PAT: Reading comprehension*. New Zealand Council for Educational Research. <https://www.nzcer.org.nz/sites/default/files/downloads/Eyre%20et%20al%202017%2C%20Mode%20equivalency%20in%20PAT.pdf>
- Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, 91(4), 483–517. <https://doi.org/10.3102/0034654321998074>
- Gath, M., Horwood, L. J., Gillon, G., McNeill, B., & Woodward, L. J. (2026). Longitudinal associations between screen time and children's language, early educational skills, and peer social functioning. *Developmental Psychology*, 62(3), 638–652. <https://doi.org/10.1037/dev0001907>
- Gowenlock, A. E., Norbury, C., & Rodd, J. M. (2024). Exposure to language in video and its impact on linguistic development in children aged 3–11: A scoping review. *Journal of Cognition*, 7(1), Article 57. <https://doi.org/10.5334/joc.385>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171–193. <https://doi.org/10.1037/amp0000307>
- Jing, M., Ye, T., Kirkorian, H. L., & Mares, M.-L. (2023). Screen media exposure and young children's vocabulary learning and development: A meta-analysis. *Child Development*, 94(5), 1398–1418. <https://doi.org/10.1111/cdev.13927>
- Kaushanskaya, M., Gross, M., & Buac, M. (2013). Gender differences in child word learning. *Learning and Individual Differences*, 27, 82–89. <https://doi.org/10.1016/j.lindif.2013.07.002>
- Law, J., Rush, R., Parsons, S., & Schoon, I. (2013). The relationship between gender, receptive vocabulary, and literacy from school entry through to adulthood. *International Journal of Speech-Language Pathology*, 15(4), 407–415. <https://doi.org/10.3109/17549507.2012.721897>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136(6), 1123–1135. <https://doi.org/10.1037/a0021276>
- Logan, S., & Johnston, R. S. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading*, 32(2), 199–214. <https://doi.org/10.1111/j.1467-9817.2008.01389.x>
- Luxon, C., & Stanford, E. (2024, August 4). *Government transforms maths education*. Beehive.govt.nz. <https://www.beehive.govt.nz/release/government-transforms-maths-education>

- May, S., Jang-Jones, A., & McGregor, A. (2019, December). *PISA 2018: Summary report—system performance and equity*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/2543/pisa-2018/pisa-2018-summary-report>
- May, S., & Medina, E. (2023, December). *PISA 2022: Aotearoa New Zealand summary report*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/PISA/pisa-2022/pisa-2022-aotearoa-new-zealand-summary-report>
- May, S., Medina, E., Ranson, C., & Fogell, D. (2023, October). *PISA 2022: Aotearoa New Zealand non-response bias analysis*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/PISA/pisa-2022/pisa-2022-non-response-bias-analysis>
- Medina, E., & McGregor, A. (2019, December). *PISA 2018: Reading in New Zealand—reading achievement and experiences of 15-year-olds*. Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/PISA/pisa-2018/pisa-2018-reading-in-new-zealand>
- Ministry of Education. (1993). *The New Zealand curriculum framework: Te anga marautanga o Aotearoa*. Learning Media. <https://natlib.govt.nz/records/21479383>
- Ministry of Education. (2007). *The New Zealand curriculum*. Learning Media. <https://newzealandcurriculum.tahurangi.education.govt.nz/new-zealand-curriculum/5637175326.p>
- Ministry of Education. (2020, December 10). *Education report: Equity Index funding decisions for Budget 2021* [Education report]. https://web-assets.education.govt.nz/s3fs-public/2024-03/16.-1245184-Education-Report-Equity-index-funding-decisions-for-Budget-2021_Redacted.pdf?VersionId=iOBxwBhxDU6yEvcFKPchT1gjYvOZS_NL
- Ministry of Education. (2022, March 28). *Implications of the literacy & communication and maths strategy. Curriculum progress tools*. <https://curriculumprogressools.education.govt.nz/support/latest-news/?ID=15>
- Ministry of Education. (2024a, July 3). *Minister outlines intentions for assessment and aromatawai*. Tāhūrangi. <https://tahurangi.education.govt.nz/assessment-and-aromatawai-announcement>
- Ministry of Education. (2024b, October 31). *National curriculum refresh reaches key milestone*. Tāhūrangi. <https://tahurangi.education.govt.nz/final-curriculum-content-october24>
- Ministry of Education. (2024c, March 1). *Te Poutāhū curriculum centre school update: 1 March 2024*. <https://www.education.govt.nz/bulletins/te-poutahu-curriculum-centre-school-update/01-03-24>
- Ministry of Education. (2025a, December 4). *Reporting to parents and whānau guidance released*. Tāhūrangi. <https://tahurangi.education.govt.nz/reporting-to-parents-and-whanau-guidance-released>
- Ministry of Education. (2025b, October). *School rolls*. Education Counts. <https://www.educationcounts.govt.nz/statistics/school-rolls>
- Ministry of Education. (2025c, October 19). *Updated curriculum content released for years 0–10 English, te reo rangatira, maths and pāngarau*. Tāhūrangi. https://tahurangi.education.govt.nz/final_y0to10_english_maths
- Ministry of Education. (2026a, July). *Attendance*. Education Counts. <https://www.educationcounts.govt.nz/statistics/attendance>
- Ministry of Education. (2026b, April). *National school roll projections*. Education Counts. <https://www.educationcounts.govt.nz/statistics/national-school-roll-projection>
- Ministry of Education. (2026c, June 11). *Phonics checks: Guidance for schools*. New Zealand Curriculum Online. <https://newzealandcurriculum.tahurangi.education.govt.nz/phonics-checks-guidance-for-schools/5637245828.p>

- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137(2), 267–296. <https://doi.org/10.1037/a0021890>
- New Zealand Council for Educational Research. (2024a). *PAT pānui | PAT reading comprehension: He aratohu hangarau | Technical guide*. <https://www.nzcer.org.nz/sites/default/files/downloads/PAT%20P%C4%81nui%20-%20Technical%20Guide%202025.pdf>
- New Zealand Council for Educational Research. (2024b, February 1). *Progressive achievement tests (PATs): What's new for 2024*. <https://www.nzcer.org.nz/news-and-blogs/progressive-achievement-tests-pats-whats-new-2024?page=6>
- New Zealand Council for Educational Research. (2025, June). *Te aroturuki kōkenga i roto i te akoranga | Monitoring progress in learning*. <https://www.nzcer.org.nz/sites/default/files/downloads/PAT%20Monitoring%20Progress%20in%20Learning%20-%20May%202025.pdf>
- Novak, J. (2023, May 15). *Approval for better start literacy approach contract variation for cohort 6* [Memorandum]. Ministry of Education. <https://web-assets.education.govt.nz/s3fs-public/2024-03/Appendix-B-v14.pdf?VersionId=gR33KU2O6DhzTljqm2hN2WWUP6nTv7dl>
- NZCER Assessment Services. (2024, July 16). *Understanding PATs for parents and the community*. <https://assessmentservices.nzcer.org.nz/support/solutions/articles/4000208374-understanding-pats-for-parents-and-the-community>
- NZCER Assessment Services. (2025, February 5). *PAT pānui | Reading comprehension: Calibration changes for 2025*. New Zealand Council for Educational Research. <https://www.nzcer.org.nz/news-and-blogs/pat-panui-reading-comprehension-calibration-changes-2025?page=4>
- NZCER Assessment Services. (2026, March 4). *Update to PAT pānui | Reading comprehension scale scores for 2026*. New Zealand Council for Educational Research. <https://www.nzcer.org.nz/news-and-blogs/update-pat-panui-reading-comprehension-scale-scores-2026?page=2>
- OECD. (2015). *Students, computers and learning: Making the connection*. OECD Publishing. <https://doi.org/10.1787/9789264239555-en>
- OECD. (2019a). *PISA 2018 results, Volume I: What students know and can do*. OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- OECD. (2019b). *PISA 2018 results, Volume II: Where all students can succeed*. OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. <https://doi.org/10.1787/062a7394-en>
- OECD, UNESCO Institute for Statistics, United Nations Children's Fund, & World Bank. (2022). *From learning recovery to education transformation: Insights and reflections from the 4th survey of national education responses to COVID-19 school closures*. UNICEF. <https://doi.org/10.1787/a79f55ac-en>
- Pacheco, E., & Melhuish, N. (2019). Exploring New Zealand children's technology access, use, skills and opportunities: Evidence from Ngā Taiohi Matihiko o Aotearoa—New Zealand Kids Online. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3461384>
- Rangahau Aotearoa Research New Zealand. (2025, August). *Aotearoa New Zealand gender attitudes survey 2025*. National Council of Women of New Zealand. https://genderequal.nz/wp-content/uploads/2025/09/NCWNZ_Gender-Attitudes-2025.pdf

- Reilly, D. (2020, July 30). Gender differences in reading, writing and language development. In *Oxford research encyclopedia of education*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.928>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress. *American Psychologist*, 74(4), 445–458. <https://doi.org/10.1037/amp0000356>
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5), 1762–1774. <https://doi.org/10.1111/j.1467-8624.2012.01805.x>
- Stanford, E. (2024a, April 29). *Education priorities focused on lifting achievement* [Press release]. New Zealand Government. <https://www.beehive.govt.nz/release/education-priorities-focused-lifting-achievement>
- Stanford, E. (2024b, May 2). *Transforming how our children learn to read* [Press release]. New Zealand Government. <https://www.beehive.govt.nz/release/transforming-how-our-children-learn-read>
- Stanford, E. (2026, May 18). *Literacy and maths at forefront of new investment* [Press release]. New Zealand Government. <https://www.beehive.govt.nz/release/literacy-and-maths-forefront-new-investment>
- Stats NZ. (2024, October 3). *2023 Census population, dwelling, and housing highlights*. <https://www.stats.govt.nz/information-releases/2023-census-population-dwelling-and-housing-highlights/>
- Taylor, G., Sala, G., Kolak, J., Gerhardstein, P., & Lingwood, J. (2024). Does adult-child co-use during digital media use improve children's learning aged 0–6 years? A systematic review with meta-analysis. *Educational Research Review*, 44, Article 100614. <https://doi.org/10.1016/j.edurev.2024.100614>
- Torppa, M., Vasalampi, K., Eklund, K., & Niemi, P. (2022). Long-term effects of the home literacy environment on reading development: Familial risk for dyslexia as a moderator. *Journal of Experimental Child Psychology*, 215, Article 105314. <https://doi.org/10.1016/j.jecp.2021.105314>
- UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education: All means all*. Author. <https://doi.org/10.54676/JJNK6989>
- UNESCO. (2023). *Global education monitoring report 2023: Technology in education: A tool on whose terms?* Author. <https://doi.org/10.54676/UZQV8501>
- von Davier, M., Kennedy, A. M., Reynolds, K. A., Fishbein, B., Khorramdel, L., Aldrich, C. E. A., Bookbinder, A., Ummugul, B., & Yin, L. (2024). *TIMSS 2023 international results in mathematics and science*. International Association for the Evaluation of Educational Achievement. <https://doi.org/10.6017/lse.tpisc.timss.rs6460>
- Watson, P. W. St J., & Rubie-Davies, C. M. (2024). Teachers' expectations of student mathematics achievement, and student mathematics self-efficacy in Aotearoa New Zealand intermediate schools: Do teacher and student gender make a difference? *New Zealand Journal of Educational Studies*, 59(1), 235–250. <https://doi.org/10.1007/s40841-023-00303-5>
- Webber, A. (2020, February). *He whakaaro: What is the relationship between attendance and attainment?* Ministry of Education. <https://www.educationcounts.govt.nz/publications/schooling/he-whakaaro-what-is-the-relationship-between-attendance-and-attainment>

