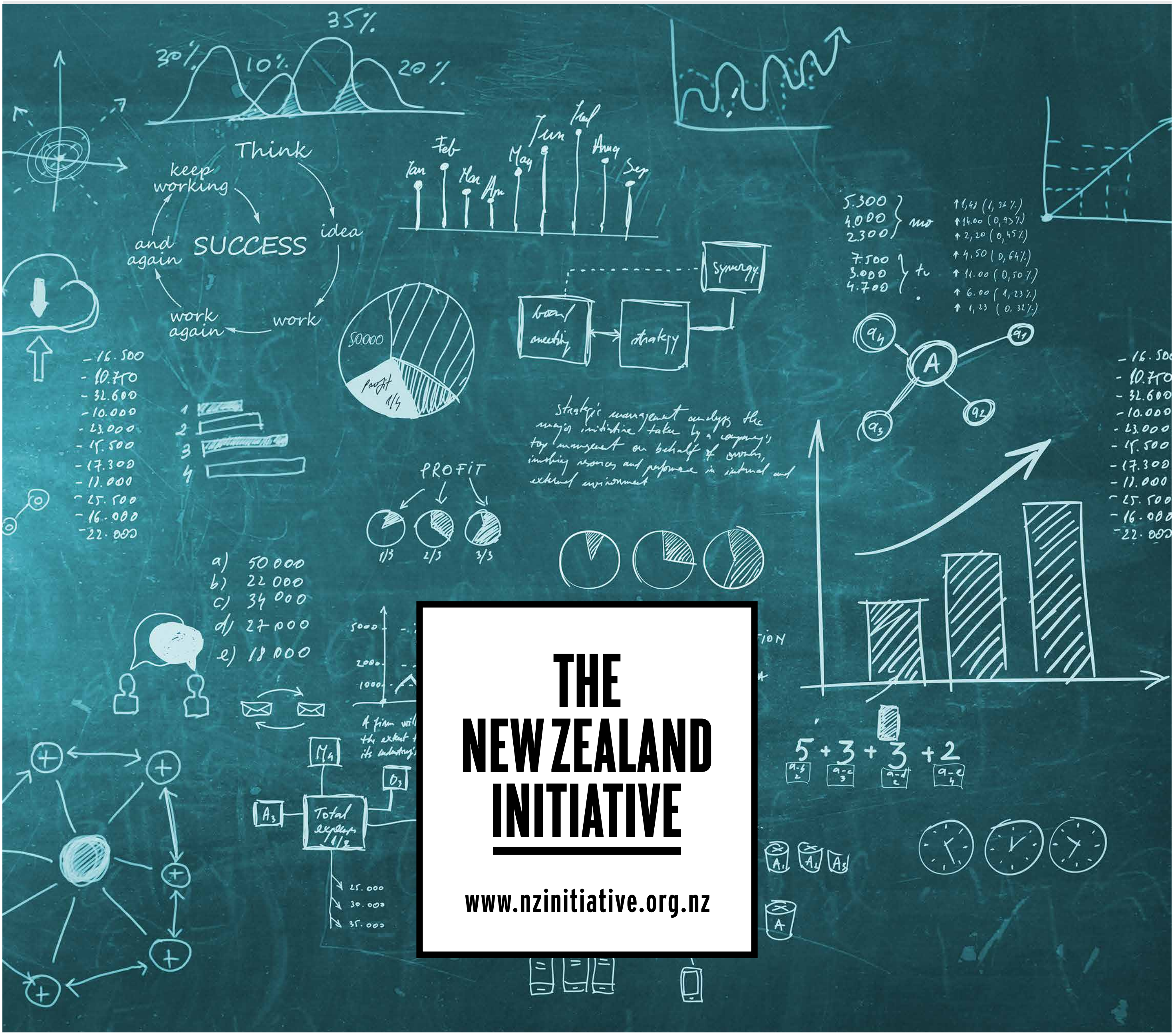
EDUCATION

SCORE!

Transforming NCEA Data

Eric Crampton and Martine Udahemuka

Foreword by Bali Haque



THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

THE
NEW ZEALAND
INITIATIVE

Published 2018 by
The New Zealand Initiative
PO Box 10147
Wellington 6143
New Zealand
www.nzinitiative.org.nz

ISBN
978-0-9951009-0-9 (print)
978-0-9951009-1-6 (online)

RR43

Designed by Angela Whitney, www.angelawhitney.com
Typeset by The Little Design Company
Cover by Mollica Sokhom

Printing arranged by True North New Zealand Ltd

SCORE!

Transforming NCEA Data

Eric Crampton and Martine Udahemuka
Foreword by Bali Haque

ABOUT THE AUTHORS

Dr Eric Crampton is Chief Economist with The New Zealand Initiative. He served as Lecturer and Senior Lecturer with the Department of Economics at the University of Canterbury from 2003 to 2014. He is co-author of several Initiative reports, including *In the Zone* and *Decade of Debt*. He holds a PhD in economics from George Mason University and authors the popular economics blog Offsetting Behaviour. He also teaches public economics at Victoria University of Wellington.

Martine Udahemuka is a Research Fellow with The New Zealand Initiative working on education policy. She is the author of *Signal Loss: What We Know About School Performance*; *Fair and Frank: Global Insights for Managing School Performance*; and *Amplifying Excellence: Promoting Transparency, Professionalism and Support in Schools.* Martine has also published on migrant and refugee integration in New Zealand. She was a student success advisor and learning consultant at Massey University, and was with the Aid Programme at the Ministry of Foreign Affairs and Trade. Martine holds a Master of Arts in Industrial and Organisational Psychology with First Class Honours from Massey University.

ACKNOWLEDGEMENTS

The authors wish to thank several people who contributed their time and expertise to this report. First, recognition goes to econometrician Rachel Hodder, formerly of The New Zealand Initiative, whose statistical work forms the basis for this report and who constructed the WRPI. Gratitude goes to the staff at the Ministry of Education for their review and to Bali Haque for his foreword. Kabir Dasgupta, Gary Hawke, Michael Johnston, Kathleen Logan, Patrick Nolan, Gail Pacheco, Grant Scobie, Audrey Tan, and Peter Winsley all gave feedback on an early draft – their input has made for a better report. The responsibility for all views expressed and any remaining errors and omissions lies with the authors.

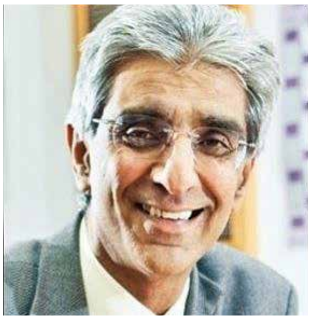
DISCLAIMER

CONTENTS

Figures

Tables

Foreword



Bali Haque

How do we know how effective our schools are? What do we know about how well our students achieve, and how good are our teachers?

These questions have bamboozled, puzzled and frustrated academics, researchers, politicians, economists, teachers, principals, parents and learners for a long time.

The problem, of course, is these are highly complex questions. School effectiveness might be measured in several ways, including collecting data about retention, truancy, suspensions, parent/student perceptions, student exit destinations, and of course qualification outcomes. Student achievement might be measured by collecting data about developing competencies and capabilities, cultural and sporting achievements, as well as qualification outcomes. Teacher quality would need to be based on some analysis of student achievement.

The authors of this report have set themselves the uphill task of critically examining one of these measures – NCEA qualification outcomes. This is entirely appropriate and timely as government moves to use 'big data' evidence, including NCEA outcomes, to guide major policy and resourcing decisions.

The Treasury, the Education Review Office (ERO), the Ministry of Education, researchers, media, tertiary institutions, employers, and many others regularly use NCEA generated data as their primary measure of the success of schools, learners and even teachers.

The authors make it abundantly clear that using NCEA data in these ways is problematic. It can lead to downstream poor decision-making and unintended impacts, which can erode the very credibility of the qualification.

A qualification without credibility soon becomes a major national problem, and we all need to be alert to this possibility. If we are going to use NCEA data for the purposes of measuring success and effectiveness, and thus to guide policymaking and resource allocation, we need a better way of doing this than we currently have.

The authors' proposed solution is a "Weighted Relative Performance Index (WRPI)," which effectively creates an overall score for each student based on the particular NCEA standards they have been assessed against. This, they argue, would enable users of the data to more credibly judge a student's success, and thus the effectiveness of schools and teachers in helping students gain valid qualifications.

The solution is elegant and highly technical, and inevitably raises questions that should be explored in future research. For example, what are the implications of creating an NCEA score in the context of a standards-based qualification? How does the WRPI fit with other possible indicators of success such as competences and capabilities? Should we attempt to measure these other indicators, and how?

There is no doubt we need more robust measures of school and student effectiveness than we currently have. The authors of this report have courageously initiated what will be a fascinating and consequential debate.

Bali Haque
Former Deputy CEO of New Zealand Qualifications Authority (NZQA)
Wellington

Introduction: Transforming education data

In 2004, 79% of students left school with at least an NCEA Level 1 certificate. By 2016, the figure had risen to 89%. But student performance on international tests has been flat or falling

Education data is great. But what you learn from it depends on what you want to see.

If you are an optimist, you will see the surge in students gaining a high school qualification since 2002 when the National Certificate of Educational Achievement (NCEA), New Zealand's main school qualification, was introduced.

In 2004, 79% of students left school with at least an NCEA Level 1 certificate – the lowest of the three NCEA Levels 1 to 3 – typically achieved around age 15. By 2016, the figure had risen to 89%.[1]

Pessimists will see the falling student performance on international tests and worry about the slipping calibre of graduates.

Realists will see the need for better measures to evaluate school performance to explain trends.

The Education Ministry and other agencies evaluating student outcomes have used attaining an NCEA qualification as the main indicator of school success.[2]

But because there are myriad ways of achieving an NCEA Level 1, 2, or 3 certificate, there can be substantial differences in performance among students with the same qualification (see Chapter 1).

Simply put, it is difficult to tell whether the surge in NCEA completion has been accompanied by more learning. Comparing and assessing performance between students in the same year, or across different years, is not a straightforward task when students take different combinations of hundreds of NCEA standards.

The New Zealand Initiative has developed a new measure of student NCEA achievement using Statistics New Zealand's Integrated Data Infrastructure (IDI) data. We describe the measure in Chapter 2 and demonstrate its applicability in Chapter 3.

No performance metric is perfect, but ours has a few advantages over other current measures. This report also illustrates the potential opportunities in the rich data warehouse available for education policy research in New Zealand (see Chapter 4). This report does not assess New Zealand's qualifications or assessment framework but rather better ways of using existing educational data.

1 Education Counts, "New Zealand Schools: Ngā Kura o Aotearoa," Website (Wellington: New Zealand Government, 2004–2016 reports).

2 The New Zealand Government reports student achievement in terms of the proportion of young New Zealanders who have achieved one of the three NCEA qualifications (NCEA 1, 2, or 3). In 2012, the Key Government set a target that by 2017, 85% of all 18-year-olds will have a NCEA Level 2 or equivalent.

IDI and better policy analysis

"At its core, SIA discards the silo mentality of social services and public agencies and uses sector-wide information and technology to understand what works in social policy, including the impact of social services, to better target investment to coordinate better policy"

Better use of large administrative datasets, or big data, has been shaping government policy dialogue.[3] This began under the Key government's data-driven Social Investment Approach (SIA), and data-driven outcomes analysis seems likely to continue under the current Labour government.

At its core, SIA discards the silo mentality of social services and public agencies and uses sector-wide information and technology to understand what works in social policy, including the impact of social services, to better target investment to coordinate better policy.[4]

The emphasis is not on inputs, money spent and saved, nor the number of people helped, but on whether the right amount of money was spent in the right areas of intervention on the right people. A dollar spent ineffectively is a dollar that could have been spent on programmes that would have done more good. Assessing outcomes matters.

Recent improvements in the collation, analysis and dissemination of microdata support this approach and help address some of the persistent policy concerns in New Zealand.[5]

The IDI links back-end administrative databases about every New Zealand resident from Census, Inland Revenue, the education system, the health care system, and more. It gives researchers anonymised information from government and non-government agencies and Statistics New Zealand surveys on education, employment, welfare, health, justice, travel and migration, and families and household. The data is securely held and can be accessed only by approved researchers under strict conditions.[6]

Education data serves research, evaluative, operational and informational purposes. It can help determine the success of national initiatives against expected outcomes.

Government and non-government analysts are already using education data in linked administrative datasets to understand connections with educational attainment and other social characteristics. For example, Treasury examined school outcomes to determine the family and individual factors that can predict underachievement.[7] The Ministry of Education (the Ministry) is considering how to use the identified risk factors to inform funding decisions for schools.[8]

3 Big data can be summarised as the application of large datasets to generate new insights about how to solve complex problems. See New Zealand Institute of Economic Research (NZIER), "Strengthening the Government's Open Investments" (Wellington: NZIER, 2015).

4 Social Investment Unit, "Briefing to the Incoming Minister" (Wellington: Ministry of Social Development, 2016).

5 Microdata refers to unit records and aggregate data collected through surveys, census or administrative processes. No information about individual people, households or businesses may be published or disseminated.

6 Statistics New Zealand, "Integrated Data Infrastructure," Website.

7 Sarah Crichton, Robert Templeton, and Sarah Tumen, "Using Integrated Administrative Data to Understand Children at Risk of Poor Outcomes as Young Adults," Analytical Paper 15/01 (Wellington: Treasury, 2015).

8 Ministry of Education, "Review of Education Funding Systems," Cabinet Paper (Wellington: New Zealand Government, 2016).

"Data on academic achievement help improve evidence-based policymaking, discover the factors associated with better academic achievement, and figure out how schools can improve student performance"

Data on academic achievement help improve evidence-based policymaking, discover the factors associated with better academic achievement, and figure out how schools can improve student performance. They also matter in other policy areas where school achievement contributes to tertiary attainment, employment outcomes, and even criminal activity. Academic achievement can be used as an input, an outcome, or a control variable to understand a number of life outcomes for individuals as well as determine the quality of education offered.

But achievement information from NCEA, which was phased in from 2002 to 2004, is not a sufficient measure for any of those objectives.

There are three levels of NCEA awarded when students achieve a given number of credits in various standards or subjects. Unlike many qualifications systems in developed countries, under NCEA, student knowledge and competency are assessed against criteria for each standard rather than against peers.

Students can build individualised academic pathways based on their interests, strengths and post-school aspirations; pursue a traditional academic pathway or can collect credits towards a vocation; choose from standards assessed by their teachers or externally; and opt from many ways of satisfying numeracy and literacy requirements.[9]

NCEA has not been uncontroversial. Opponents say it encourages 'credit farming,' where students, teachers and schools opt for courses offering easy credits towards certification. Proponents praise its flexible pathways to success, particularly for students who want to pursue non-traditional courses such as fine arts and tourism.

To illustrate, about five years after NCEA was introduced, former teacher Peter Joyce explained how students had learned to game the system:

> At a school where I taught, senior students were grinding out assessment points in tourism. They astutely figured out that it is easier to draw up an itinerary for a fly-drive holiday in Tahiti than to get to grips with the special theory of relativity.[10]

A decade on and teacher Peter Lyons illustrated what he saw as the flaw in NCEA:

> My nephew … has figured out that it is much easier to pass internal assessments than the external assessments sat as exams at the end of the year …
>
> [He] has also figured out that certain subjects and certain units within these subjects are much easier to pass than others …[11]

9 New Zealand Qualifications Authority (NZQA), "NCEA," Website.

10 Peter Joyce, "System fatally flawed," *The Christchurch Press*, cited in Muriel Newman, "How good is our education system?" (New Zealand Centre for Political Research, 4 June 2007);

11 Peter Lyons, "NCEA pass rates hard for parents to cope with," *The New Zealand Herald* (25 November 2015).

In response, the Ministry's head of student achievement said:

> NCEA is not a pass or fail system. Our National Certificates of Education Achievement allow kids to learn and achieve as they go. While students sit exams, there's much more to it…
>
> …
>
> But crucially, where a student has not fully met the standards for a credit during the year, they can have another shot after further teaching and learning.
>
> …
>
> As well as the traditional subjects, NCEA allows kids to develop a diverse range of highly-relevant skills and knowledge. And that's exactly what employers are looking for.[12]

"Given the varied ways NCEA can be attained, simply noting whether the highest attained qualification is Level 1, 2 or 3 is less informative than it could be"

New Zealanders need to be confident in the value of educational qualifications. The New Zealand Initiative's 2018 report, "Spoiled by Choice: How NCEA hampers education, and what it needs to succeed," analysed the merits and shortcomings of the NCEA framework.[13]

But, the intention in this current report is not to discuss New Zealand's qualifications framework. Instead, we argue that current NCEA achievement information impede meaningful judgments about student ability.

What is inside the qualification matters, but the system is geared towards counting attained qualifications rather than weighting their content. That may be desirable if the purpose of NCEA is to recognise achievement across disparate courses; instead, NCEA makes it tough to assess academic ability, and presents schools with incentives to get students through the path of least resistance rather than the path most suited to their abilities.

Given the varied ways NCEA can be attained, simply noting whether the highest attained qualification is Level 1, 2 or 3 is less informative than it could be.

It is a problem in research using these levels as a student performance metric.[14] The challenges extend to tertiary education providers and employers who wish to understand relative competency or ability when comparing two students with the same qualification.

Finer measures are needed where performance variation within certificates is high.

Alternative and finer measures exist and have merit, but the measure proposed in this report is preferable for some purposes and a useful complement in a suite of measures.

We use our new measure in a case study in association with a previous Initiative report on numeracy, "Un(ac)countable: Why Millions on Maths

12 Lisa Rodgers, "NCEA is not a pass or fail system," *The New Zealand Herald* (26 November 2015).

13 Briar Lipson, "Spoiled by Choice: How NCEA hampers education, and what it needs to succeed" (Wellington: The New Zealand Initiative, 2018).

14 Looking at the list of research projects on education using the IDI database on the Statistics New Zealand website, educational outcomes in many are defined in terms of the overall NCEA certificate gained. Statistics New Zealand, "How researchers are using the IDI: Education and training," Website.

"What is inside the qualification matters, but the system is geared towards counting attained qualifications rather than weighting their content"

Returned Little," which concluded that declining numeracy among primary school teachers may be partially to blame for declining student numeracy scores in international rankings.[15] We tested this hypothesis in this report and found that, as best we can measure it, numeracy among those pursuing tertiary degrees in primary school teaching has been flat or increasing, so declining teacher numeracy is unlikely to explain declining student maths test scores.

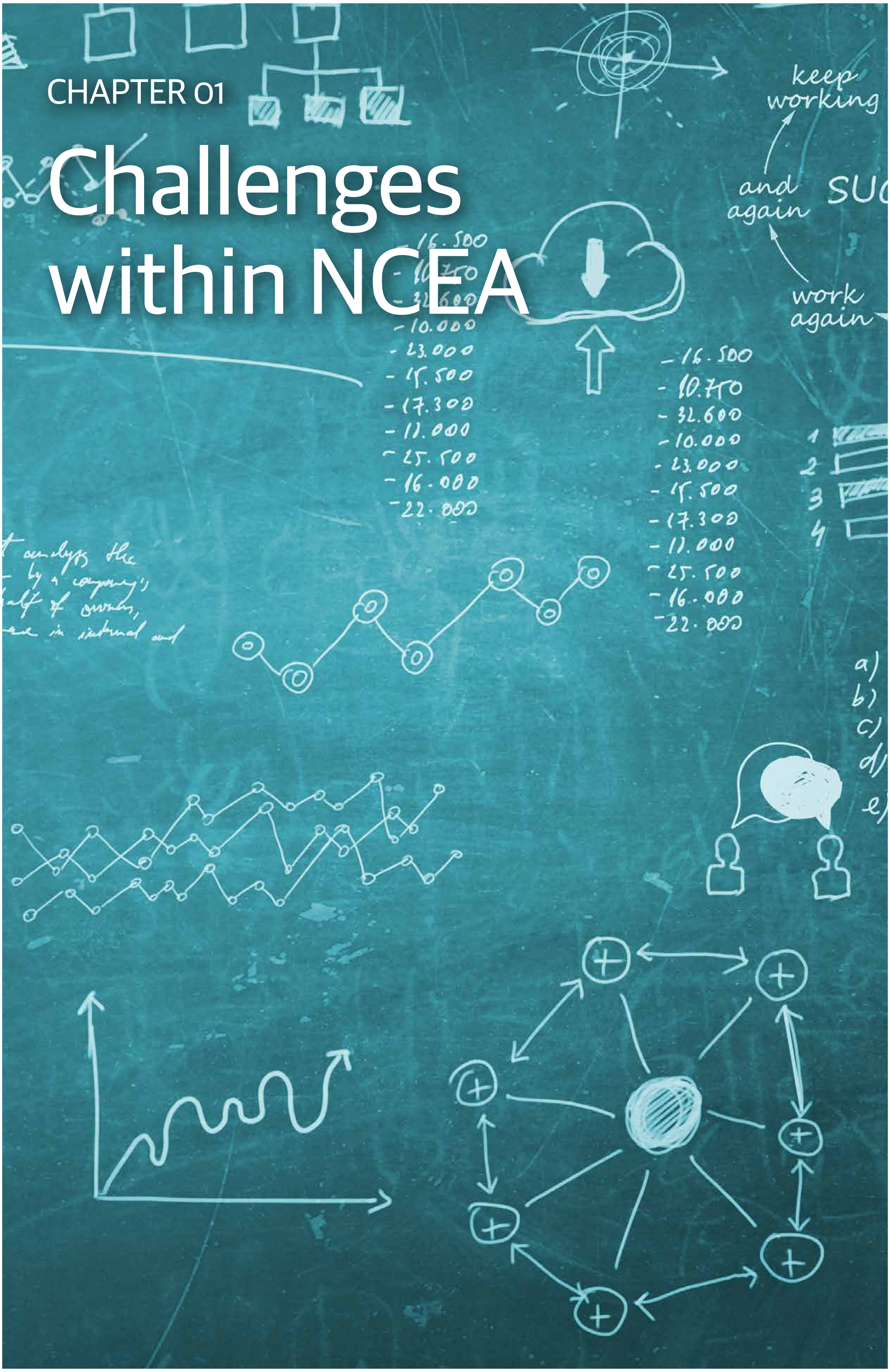
Better achievement measures can help students realistically assess their own performance, and employers and universities in selecting candidates, the Ministry in assessing school performance and the effects of policy changes, and research to explain the determinants and consequences of educational performance.[16]

15 Rose Patterson, "Un(ac)countable: Why Millions on Maths Returned Little" (Wellington: The New Zealand Initiative, 2015).

16 Of course, the metric would form part of a suite of other variables to be considered for these uses. To determine the quality of schools, for example, would require knowing more about the students the school serves. The New Zealand Initiative has argued for better measures of school performance in past reports. (For an account of other out-of-school factors known to influence school achievement, see Martine Udahemuka, "Amplifying Excellence" (Wellington: The New Zealand Initiative 2017)). Having a metric that provides more meaningful information about achievement makes this work easier.

CHAPTER 01

Challenges within NCEA



"However, because equivalent NCEA qualifications can be achieved via radically different pathways, comparing student performance is more challenging"

NCEA is radically different from both New Zealand's previous and many international qualifications frameworks. The standards-based and flexible framework which aims to cater for the diverse needs and strengths of students has transformed the education landscape in New Zealand.

However, the in-built flexibility is both a strength and weakness of NCEA. The flexibility lets teachers and schools customise their offerings to the needs of students. However, because equivalent NCEA qualifications can be achieved via radically different pathways, comparing student performance is more challenging.

Even within a given subject, it can be difficult to interpret NCEA results as a meaningful indication of mastery of the domain. Instead of an overall grade for maths, students are awarded separate grades for standards in trigonometry, algebra, geometry, etc. This can be an advantage if an employer needs to know whether a student understands trigonometry or algebra, for example. But it can be less informative about overall ability in maths than a comprehensive score across a broad range of math subjects.

Students generally sit NCEA Levels 1 to 3 in Years 11 to 13. However, they can sit standards outside their year level. For example, Year 13 students can sit NCEA Level 2. This must be considered when constructing a student performance measure.[17]

NCEA assessment standards

Students can sit 'unit' and 'achievement' standards, each worth a different number of credits that count towards a total required to achieve an NCEA certificate at a given level.

Unit standards are all internally assessed and have learning elements, all of which must be met. Unit standards are usually graded as pass/fail:

- Achieved (A) for meeting the criteria of the standard
- Not achieved (N) if a student does not meet the criteria of the standard[18]

Achievement standards can be internally or externally assessed, and have a single criterion that can be graded as:

- Achieved (A) for a satisfactory performance
- Merit (M) for very good performance
- Excellence (E) for outstanding performance
- Not achieved (N) if students do not meet the criteria of the standard[19]

The discrete nature of the four-category grade system adds another level of difficulty when producing a student performance metric. While the difference between 77% and 78% in a university course is likely statistical noise, four performance categories are rather coarse.

17 Though it may be more common for students to be doing catch-up standards from a lower standard for their year group than for students to do accelerated standards from a higher level.

18 New Zealand Qualifications Authority (NZQA), "Standards," Website.

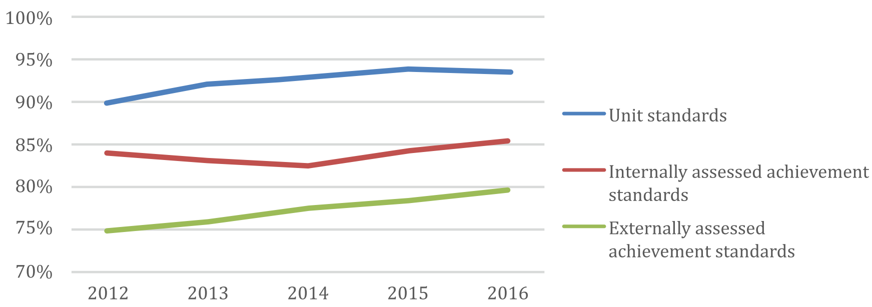
19 Ibid.

"Passes in different standards also mean different things"

Passes in different standards also mean different things. Whereas almost 95% students pass unit standards and about 85% pass internally assessed standards, pass rates for externally assessed standards are just under 80% (see Figure 1). Since 2012, the average percentage of students receiving a passing grade in unit standards has been above 90%.

Comparing student performance across different years is also complicated within NCEA. Pass rates overall are higher now than in 2012 and while student performance may have improved, it is possible that grading standards have also changed – or that students have deduced which standards they are most likely to pass.

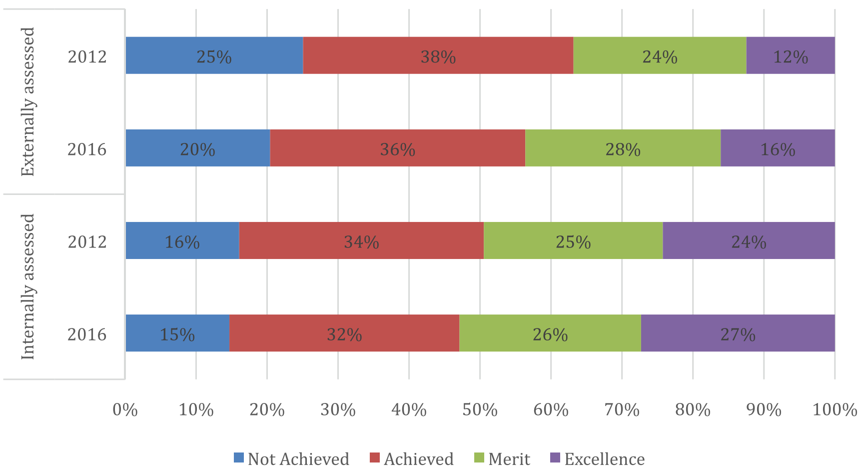
Figure 1: Average NCEA pass rates by assessment type (2012–16)



Source: New Zealand Qualifications Authority (NZQA), "Secondary Statistics Consolidated Files," Website.
Note: Data collated from Standard Achievement Statistics files for all available years.

The difference in grade distributions between internally and externally assessed achievement standards is starker in Figure 2. In 2016, more than 50% students in internally assessed standards received a *merit* or *excellence* (compared to 44% in externally assessed). Consequently, an NCEA *achieved* now means, or may soon mean, performance worse than half of the students sitting the standard.

Figure 2: Average NCEA grade distribution for achievement standards by assessment type (2012 and 2016)



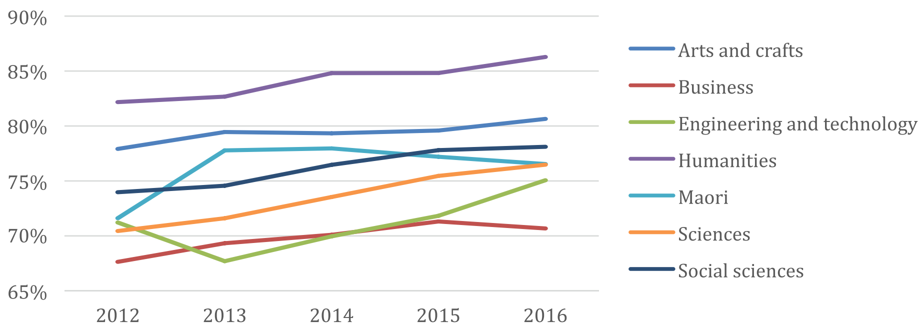
Source: New Zealand Qualifications Authority (NZQA), "Secondary Statistics Consolidated Files," Website.
Note: Data collated from Standard Achievement Statistics files for all available years (observations: 27,981).

NCEA subject differences

"Pass rates in business subjects are at least 10 percentage points lower than in humanities in each of the years surveyed"

Substantial differences in pass rates across subjects between 2012 and 2016 are clear in Figure 3. Pass rates in business subjects are at least 10 percentage points lower than in humanities in each of the years surveyed. Pass rates have been increasing in most subjects, implying that the general increase in NCEA attainment is not driven by any particular subject area.

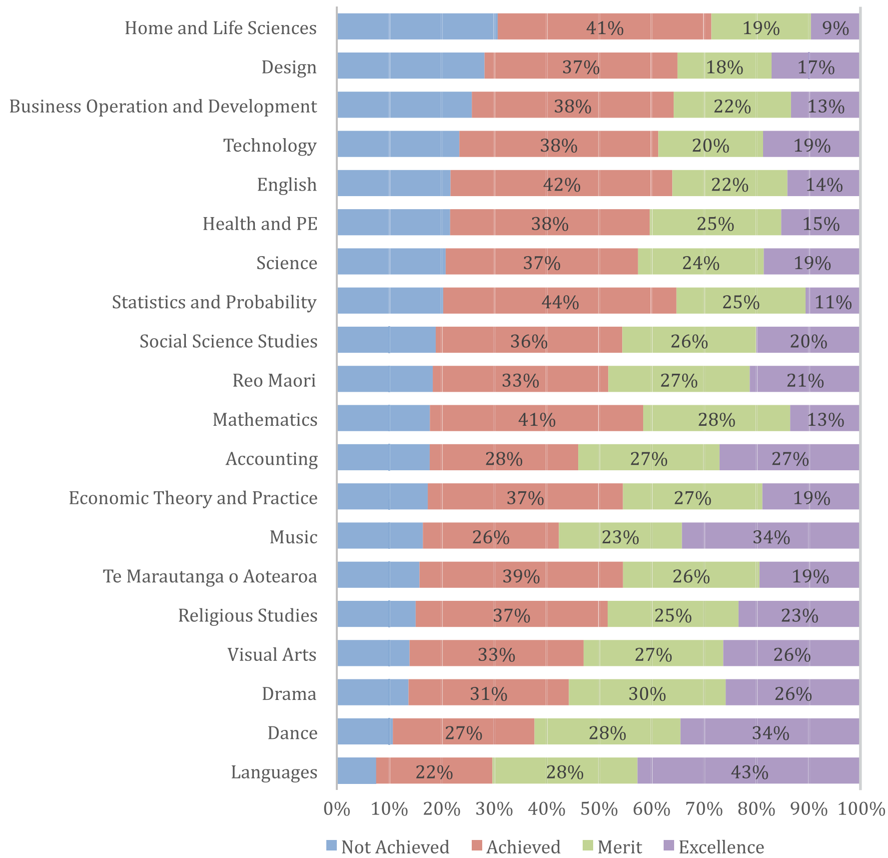
Figure 3: Average NCEA pass rates by subject field (2012–16)



Source: New Zealand Qualifications Authority (NZQA), "Secondary Statistics Consolidated Files," Website.
Note: Data collated from Standard Achievement Statistics files for all available years.

As shown in Figure 4, there is clear variation in the grade distribution across subject sub-fields. Students choosing courses that suit their abilities may explain some of the variance across fields. But it would be difficult to compare an excellence grade in Home and Life Sciences with one in Languages, for example.

Figure 4: Average NCEA grade distribution for achievement standards by subject subfield (2012–16)



Source: New Zealand Qualifications Authority (NZQA), "Secondary Statistics Consolidated Files," Website.
Note: Data collated from Standard Achievement Statistics files for all available years (observations: 6,365).

Current student performance metrics

The bluntest metric of student attainment in New Zealand is the NCEA qualification level.

NCEA Level as an indicator of performance

"Failing to attain NCEA Level 2 may signal something about a student, but it is difficult to tell what achieving a Level 2 means without looking closely at a student's transcript"

The Key Government's goal of "boosting skills and employment" under Better Public Service (BPS) targets sought to increase the proportion of 18-year-olds with NCEA Level 2 or equivalent.[20] Inevitably then schools have been judged, in part, on whether at least 85% of their students leave with that qualification. Given an NCEA certificate may be packaged and gained in hundreds of different ways, simply noting the NCEA level achieved masks vast variations in student ability – and in the quality of the qualification. It also makes it difficult to assess the quality of the schooling system. The blanket Level 2 BPS target may thus encourage schools to push students into easier rather than challenging standards, which means year-on-year changes in attainment rates may not be meaningful.

NCEA levels have also been taken as a key variable of education attainment in several studies using IDI data.[21] But where variation within the qualifications is high, NCEA attainment provides only a crude measure of performance. Failing to attain NCEA Level 2 may signal something about a student, but it is difficult to tell what achieving a Level 2 means without looking closely at a student's transcript.

But the NCEA certificate level is only the bluntest of available measures or indicators of achievement as it is essentially a measure of quantity (number of credits collected) rather than a measure of quality.

NCEA-derived metrics

NCEA student performance information has been converted in other ways to derive more meaning about student ability. Besides the NCEA level, the commonly used measures in the sector are: total credits achieved, total credits achieved at *merit* or *excellence*, Cumulative Score, and Expected Percentile.[22]

The first two provide more information about student performance than by simply observing whether a student obtained a given NCEA certificate. Both approaches can be built on to construct better measures. The last two measures provide reasonably minute information on student performance.

20 State Services Commission, "Better Public Services 2012–2017," Website.

21 See, for example, the list of research projects in education using the IDI database, many of which define educational outcomes in terms of the overall NCEA certificate gained. Statistics New Zealand, "How researchers are using the IDI,"op cit.

22 For more on these metrics, see Scott Ussher, "Post-School Choices: How Well Does Academic Achievement Predict the Tertiary Education Choices of School Leavers?" (Wellington: Education Counts, 2008), 22.

The Cumulative Score assigns a point value to different grades for each achievement standard sat:

- Excellence: 4 points
- Merit: 3 points
- Achieved: 2 points
- Not achieved: 0 points

This value is multiplied by the number of credits on the standard and then tallied across all achievement standards taken by a student at each level. The Cumulative Score omits performance on unit standards in its calculations as those are assessed on an *achieved/not achieved* basis.

There is no particular basis for the points awarded to each achievement level in the cumulative score. And schools use different weightings in their own top academic Dux awards to students.[23] Recent analysis by Tim Maloney and Kamakshi Singh shows that the cumulative score undervalues *merit* scores relative to *excellences*, but that *achieved* credits on their own do little to predict future success.[24] But giving points for the number of *excellences* can be a problem if they are easier to achieve in some subjects than in others.[25]

The Expected Percentile metric adjusts somewhat for the difficulty of each standard. It converts each student's score in each standard into an expected percentile ranking for that standard. It then calculates an average of these estimates over all standards taken by a student to give an 'expected percentile' at each NCEA level.

For example, if a student earns an *excellence* in a standard where only 10% students earn one, that student's performance is at the 95th percentile on the Expected Percentile. The performance is higher than the 90% students who did not attain an *excellence*, and is expectationally at the midpoint of those awarded an *excellence* – 90% plus half of 10% is 95%. How this works for a single standard is illustrated in Table 1. Students who did not achieve the sample standard are in the 10th percentile because 20% students also received a *not achieved* grade (10 being the midpoint between 0 and 20).

23 See, for example, Rathkeale School ("Information for Students, Parents, Teachers: Criteria to Determine Dux and Proxime Accessit, 2015) and Faircol School (Process for Calculation of Dux and First Place in Years 11 and 12) – from a random selection of schools in New Zealand.

24 Tim Maloney and Kamakshi Singh, "Using Validated Measures of High School Academic Achievement to Predict University Success," AUT School of Economics *Working Paper Series*, 2017/10 (AUT School of Economics, December 2017).

25 Similarly, Meehan et al (2018) find that differences in NCEA Level 1 performance substantially explain ethnic differences in enrolment in Bachelor's-level study, with differences in the number of merit and excellence credits being most important. Adjusting for the difficulty of attempted credits could explain a greater proportion of the variation if students from different backgrounds choose different courses. Lisa Meehan, Gail Pacheco, and Zoe Pushon, "Explaining ethnic disparities in bachelor's degree participation: evidence from NZ" (Studies in Higher Education, 2018).

Table 1: Results distribution of a sample achievement standard using the Expected Percentile measure

Result	Frequency A	Percentage B	Cumulative percentage C	Expected percentile $D_i = (C_{i+1} + C_i)/2$
Excellence	12	10%	100%	95%
Merit	24	20%	90%	80%
Achieved	60	50%	70%	45%
Not achieved	24	20%	20%	10%

Source: Education Counts, "Post-school Choices: How Well Does Academic Achievement Predict the Tertiary Education Choices of School Leavers?" (Wellington: Ministry of Education, 2008), 22.
Note: This sample standard is an externally assessed achievement standard with 120 results.

The following chapter discusses the strengths and weakness of existing performance measures and desirable properties of any performance index. We then present an alternative metric, the Weighted Relative Performance Index (WRPI), which overcomes some of the limitations in existing measures.

CHAPTER 02

Weighted Relative Performance Index (WRPI)

To construct a performance measure, the user needs to know what they want from it. A student performance metric can be used for research, to inform students of their performance, and possibly for operational use.

We have developed a new measure, the Weighted Relative Performance Index (WRPI), as one way of supporting these goals.

Desirable properties in performance metrics

Six desirable properties in a student performance metric are that the measure:

1. is easily comparable between students;
2. provides meaningful levels of gradation;
3. considers competence at different levels;
4. adjusts for difficulty in assessments;
5. is not overly influenced by results in a single assessment; and
6. does not encourage gaming of the measure.

A valuable measure would quantify performance and be easily comparable across students. But the quantification should only be reported at the finest level for which precision is possible. A measure that suggests huge differences between one student earning 78% and another earning 79.5% on the same test is probably not that reliable – but neither is one that lumps a student earning 50% with one earning 75%.

A measure should consider the grade given to students. Those receiving *excellence* should be assigned higher values than those receiving *merit* or lower in the same NCEA standard. If two students take the same set of standards, and one student never receives a worse score and sometimes outperforms the other, the stronger student should receive a higher score on the measure.

A measure should also account for assessment difficulties across subjects and years. *Excellence* in some standards might be easier to achieve than a *merit* elsewhere.

NCEA pass rates vary greatly across subjects, unit and assessment type, and years (see Chapter 1); thus, a good performance metric should consider how different standards have starkly different grade distributions. More students pass unit standards and internal assessments than external tests (see Figure 1).

A student who achieves standards demonstrably harder to pass should be assigned a higher value on a performance metric than one who achieves identical results but in standards easier to pass. Likewise, students who achieve high grades in years in which fewer students were given high grades should be given a higher value on the metric.

Additionally, a metric should not be unduly influenced by the result from a single NCEA standard. It is possible for a student to 'fluke' a much better result than they would normally get or have unfortunate circumstances that affect their grade on one standard. An accurate metric would not be overly sensitive to a single result.

"Better measures would weigh both the quantity of credits and the quality of those credits, and not discourage students from taking the courses that are best for them"

Student, teacher or school choices are also affected if a metric is used operationally, is used to inform, or has a reward attached. Targeting NCEA completion rates, for example, risks encouraging schools to shepherd risky students through easy pathways. Better measures would weigh both the quantity of credits and the quality of those credits, and not discourage students from taking the courses that are best for them.[26]

We argue here for an additive measure. Imagine two students taking an identical set of courses and earning identical grades. If one of those students then attempts an additional, more challenging standard, and earns a lower grade than on the other standards, should that student be ranked below the student who never attempted the more challenging course?

Averaging performance across subjects into a single score can discourage students from sitting standards where relatively high performance is more difficult. For example, if the best students – comprising 40% of the cohort – take calculus and all attain an *excellence*, the very best student could never attain higher than the 80th percentile in that course using the Expected Percentile ranking, and a middling student receiving only an *achieved* might be lucky to reach the 40th percentile.[27]

Adding scores cumulatively rather than averaging them can solve this problem, but risks masking other important information revealed when a student *fails* a standard. The implication is that incentive problems are more important than the loss of information in ignoring failed standards.

Constructing a performance index

Many desirable properties overlap between a student performance index and other indices.

Typical microeconomics courses deal extensively with functions that represent the conversion of various inputs into a single output. These functions are most commonly used to demonstrate how firms convert combinations of labour, capital and land into consumption goods, or how consumers can convert bundles of goods and services into wellbeing.[28] The methods used in constructing the indices are generalisable beyond problems of production and consumption.

The Cobb-Douglas function is frequently used to represent converting multiple inputs to a single output. Economists like it as it reflects monotonicity and convexity – staple assumptions in representing

26 One referee pointed out the need to consider differences in the usefulness of different standards. If attaining an *excellence* in Latin is tough but does little to improve future life outcomes than an easier-to-achieve *excellence* in calculus, WRPI could distort choices. WRPI can easily be disaggregated to subject-level scores (such as languages, maths, business), and those subject scores could be used to predict life outcomes. Important differences across sub-indices could be revealed and considered in future research.

27 One of the authors of this report, in a prior life as senior lecturer in economics at Canterbury, despaired at top students who avoided more difficult essay-based courses to preserve their scholarship-attracting A+ GPA. The measure distorted those students' choices: penny-wise but pound-foolish.

28 See, for example, Hal Varian, *Microeconomic Analysis*, 3rd ed. (Norton, 1992), 4.

consumption preferences and production ability. These properties are also appropriate for evaluating student performance.

Monotonic loosely means 'more is better' – or in the case of NCEA, the more standards passed, the higher the index value. The index value cannot fall for students taking extra standards. Likewise, a student with better performance cannot achieve a lower value on the metric.

Convexity refers to averages being weighted relatively higher than extremes. For this purpose, it means a student with decent results across the board will generally outperform a student who achieves a few good results alongside mostly poor results. The impact on the index value for a good result in a single standard should not be excessive.

Hence, an index styled to mimic a Cobb-Douglas function can incorporate the desirable properties of a student performance index. Our index is then:

$$WRPI_j = \sum_{i=1}^{n} \propto_i \ln_i x_{i,j}$$

where $WRPI_j$ gives the WRPI index score for student j; $\propto_i$ gives the number of credits for standard i; and $x_{i,j}$ denotes the relative performance on that standard as shown by the inverse proportion of students who achieved the same result or better than student j.

$$x_{i,j} = \frac{(\textit{no. of students who sat standard } i\,)}{(\textit{no. of students who received the same or better grade than student } j \textit{ on standard } i)}$$

Very simply, where the expected percentile ranking takes an average of percentile scores across all courses attempted, our metric adds those percentile scores together using a log weighting, while also making it possible to add in scores from unit standards evaluated as pass/fail.

The WRPI properties

This section discusses how current metrics and the proposed WRPI metric stack up against the properties outlined above.

A scenario of six students sitting various combinations of five possible standards is presented in Table 2. Standards 1–4 are achievement standards with differing grade distributions. These reflect the likely relative difficulty of the assessment. Standard 1 has 64% students failing the standard, and 1% receiving *excellence*. In contrast, 20% students gained *excellence* in Standard 3. Standard 5 is a unit standard with a 90% pass rate.

The bottom half of Table 2 presents six students, labelled A through F, in a plausible rank ordering by performance. Reasonable people may disagree about the relative positioning of students C and D, but it would be difficult to argue that Student A did not outperform Student B. While different metrics might produce different rankings of students C and D, but should not reverse the rankings of Students A and B.

Table 2: Hypothetical sample of student grades

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Standard type	**Achievement standard**	**Achievement standard**	**Achievement standard**	**Achievement standard**	**Unit standard**
# of credits	2	3	4	3	3
% excellence	1%	10%	20%	5%	–
% merit	5%	30%	25%	10%	–
% achieved	30%	40%	40%	50%	90%
% Not achieved	64%	20%	15%	35%	10%
			Grades		
Student A	E	E	E	E	A
Student B	M	E	-	M	A
Student C	A	M	E	NA	–
Student D	A	A	M	A	A
Student E	NA	A	A	NA	A
Student F	–	A	A	–	NA

Note: E: Excellence; M: Merit; A: Achieved; NA: Not Achieved

Table 3 shows how students would be scored on the various metrics, including WRPI.

The choice of metric can have significant effects on the relative rankings of students; the case here was chosen to illustrate these differences. In some cases, the discrepancy in rankings could be a problem depending on the purpose for using the metric.

Table 3: Results for students from Table 2 by metric

	Total achieved credits	Total excellence credits	Cumulative score	Expected percentile	WRPI
Student A	15	12	48	95.5%	31.9
Student B	11	3	27	93.8%	18.5
Student C	9	4	29	65.4%	11.2
Student D	15	0	28	61.6%	7.5
Student E	10	0	14	31.1%	1.6
Student F	7	0	14	37.5%	1.3

From the first column in Table 3, simply summing up the credits achieved is a poor indicator of performance. Students A's and D's performance is sharply different despite having the same number of achieved credits. NCEA attainment is effectively a metric of total achieved credits.[29]

[29] The requirement to achieve an NCEA certificate is a given number of credits at each level. The only exception is Level 1, which imposes additional numeracy and literacy requirements. New Zealand Qualifications Authority (NZQA), "NCEA levels and certificates," Website.

Moving across in the table, adding the 'quality' of student performance by calculating the number of credits received with *excellence* can too yield poor rankings. Not only does Student C become higher ranked than Student B but also the stark variation in performance between Students D and Students E and F is completely masked.

Endorsements for an NCEA certificate effectively reflect this type of metric. They are given to students who achieve at least 50 credits at *merit* or *excellence* at the level of the certificate or above.[30] However, it is entirely possible for students who failed to reach the threshold for an endorsement to have outperformed those who did meet the threshold, depending on the level of difficulty of the standards taken.

From the third column, the cumulative score better reflects performance variation but still produces obscure rankings. The cumulative score ranks Students C and D above Student B, but Student B outperformed the others in all the standards that all three students sat.

The strength of the cumulative score is it differentiates the level of performance and recognises that an *achieved* grade demonstrates lower performance than an *excellence*. But as it does not adjust for difficulty between standards, it does not provide a strong way of assessing students' relative abilities. It also masks the difference between Students E and F as it does not consider unit standards. Were it to incorporate unit standards by awarding two points for a pass, it would rank Student D ahead of Student B despite Student B outperforming (or tying) Student D on all standards both sat.

Although a much more sophisticated measure of performance than the preceding three metrics, Expected Percentile also has shortcomings. Student F is ranked higher than Student E, despite the latter doing no better than Student F in any standard. The reason is it treats a failed attempt at a standard differently to a no attempt. Since the expected percentile takes an average rather than a sum, including a failed standard can quite drastically affect the value on the metric.

For a similar reason, Student A is ranked lower than Student B despite Student A achieving perfect results in all standards. In this case, Student A's score on the metric is brought down by including the relatively easy Standard 3. Student B did not try the standard and so did not have their value on the metric brought down by the one relatively easy standard. Not only did Student A sit more standards than Student B, but Student A also outperformed or matched Student B's grade on all standards – even so, Expected Percentile ranks Student A lower than Student B.

In the final column, the WRPI does not suffer the same issues apparent in the other metrics. It preserves the intuitive ranking of students, is not overly sensitive to a student's performance on a single standard, and does not 'punish' students for taking additional standards.

However, it could be vulnerable to a perverse rank ordering if, for example, asked to compare one student who earned 80 credits with *excellence* to another who earned 320 credits with *achieved*. The former would almost certainly be the better student. We thank one of our referees for reminding us of this point, but note that it would be difficult to use

[30] New Zealand Qualifications Authority (NZQA), "NCEA endorsements: Recognising high achievement with 'endorsements'," Website.

this kind of 'brute force' attack to game WRPI: accumulating too many credits costs performance in every other attempted credit because time is spread too thinly.

"WRPI could be of both academic and operational use. Achieving a high score on WRPI requires passing more standards and with higher grades. The index would not influence the trade-off between time studying for more or fewer standards. That decision can be left to educators and students"

WRPI could be of both academic and operational use. Achieving a high score on WRPI requires passing more standards and with higher grades. The index would not influence the trade-off between time studying for more or fewer standards. That decision can be left to educators and students.

Importantly, it would encourage students to sit challenging rather than easy standards. A student trying to earn the highest score with the least effort would aim for those courses where the student is likely to excel relative to other students, rather than simply aiming for the courses that are easiest to pass.

Hence, if this metric is published on student records or used to evaluate school performance, it is less likely to lead to gaming than a metric that punishes students for taking additional standards.

One important deficiency remains. If students of different ability select different courses, WRPI and other measures would mask very real differences in performance.

Consider economics and sociology. In each course, suppose equal proportions of students earn *excellence*, *merit*, *achieved* and *not achieved*. Suppose further that every student who passes sociology with *excellence* earned only a *not achieved* or *achieved* in economics, and that no student earning higher than an *achieved* in economics would consider taking sociology. Both courses would be considered similarly difficult under both the expected percentile measure and under WRPI, but the economics paper is clearly the more difficult one in this scenario.

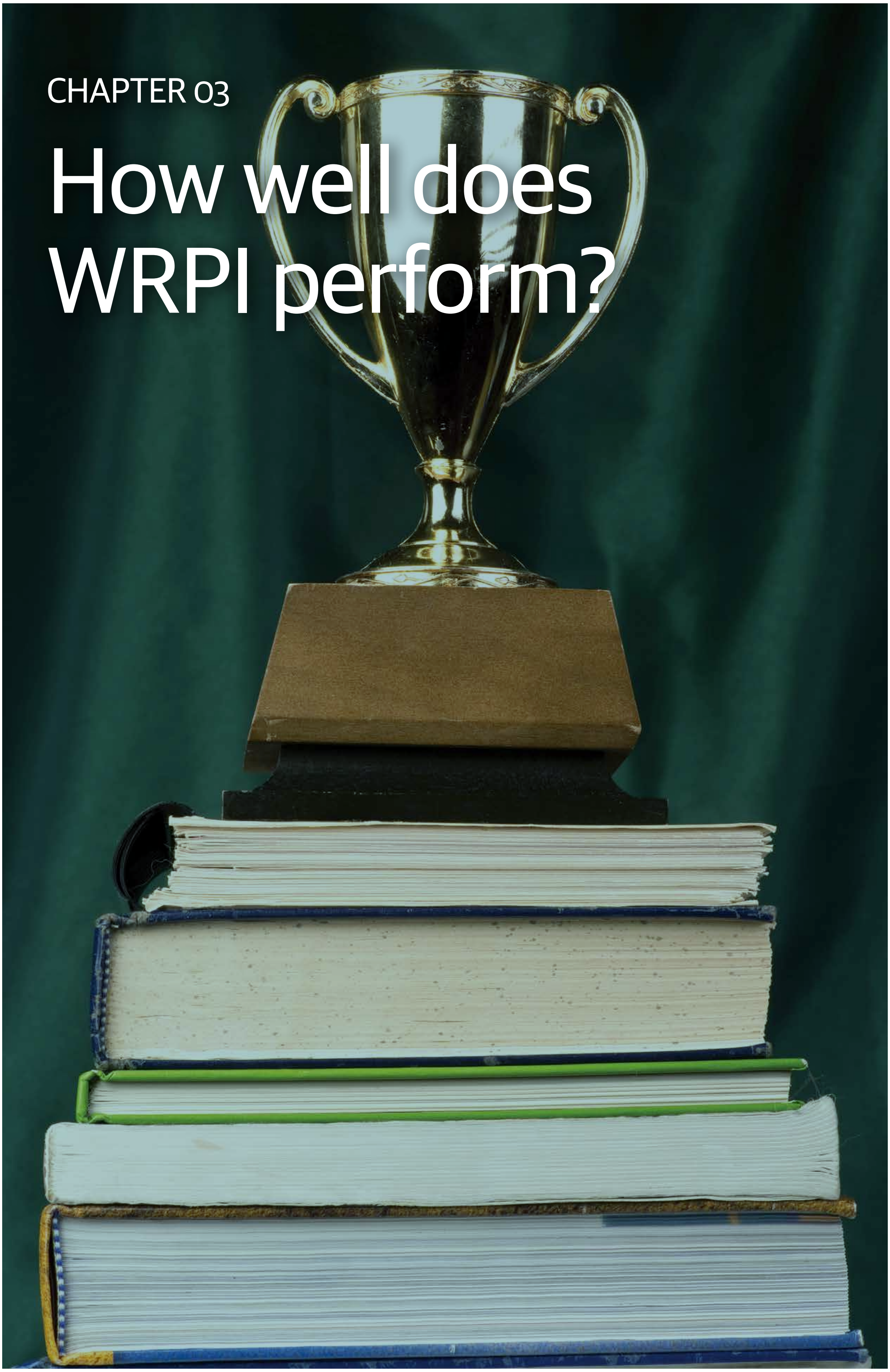
A more comprehensive ranking tool would weigh the relative difficulty of the courses to account for this – and prove computationally intensive.

That task is beyond the scope of this report.[31]

[31] The University of Canterbury produced a Difficulty Index of its courses ranking a course's difficulty based on how well students performed in that course, on average, compared to their performance in other courses. So if everyone taking intermediate macroeconomics earned, on average, a letter grade lower score than in other courses, intermediate macroeconomics would be rated as relatively difficult. Producing that measure across the thousands of NCEA standards would be computationally intensive.

CHAPTER 03

How well does WRPI perform?



The student scenario presented in Chapter 2 demonstrates that WRPI can perform well in circumstances where other metrics struggle.

But one example constructed to highlight differences in how different measures would rank students proves little. We went to Statistics New Zealand's Integrated Data Infrastructure (IDI) to check WRPI's results and apply the measure to a large set of students who had sat NCEA.

If WRPI correlates well with other broad measures of student performance, that lends confidence to its concurrent reliability. If it also predicts student progression to tertiary study, completion of tertiary study, and later earnings, the measure would prove useful for a broad range of other academic and policy-oriented work.

For example, if increases in WRPI are associated with higher tertiary attainment and earnings, the government might use those figures to weigh programmes that improve student attainment. Future research could look at whether disaggregated versions of WRPI scoring students in different subjects add predictive power. For example, WRPI scores in STEM (Science, Technology, Engineering and Maths) subjects might do more to predict future earnings than scores in other subjects.

Ideally, WRPI should be run against other available measures, like the Cumulative Score and Expected Percentile, to see which measure best predicts post-school outcomes. Time constraints prevented our IDI analyst from undertaking the 'horse race' of our measure against competing alternatives in regression analysis, so anyone wanting to use WRPI should do this stage first.

To check how WRPI performs, we first correlated the measure at each NCEA level, and a maths-specific WRPI measure, with the other performance measures, including:

- the number of credits at each level completed by age 18;
- number of credits received with excellence at each level; and
- Cumulative Score at each level; and
- Expected Percentile at each level.

Future research could distinguish between enrolment at university and polytechnics.

The results are presented in Table 4.

Table 4: Correlations between WRPI, tertiary enrolment, and other measures for all students born between 1988 and 1996

	Enrolled in a Tertiary qualification before 2016	Number of level 1 credits completed by age 18	Number of level 2 credits completed by age 18	Number of level 3 credits completed by age 18	expected percentile for level 1	expected percentile for level 2	expected percentile for level 3	WRPI score for level 1 – all subjects	WRPI score for level 2 – all subjects	WRPI score for level 3 – all subjects	WRPI score for level 1 standards in either the maths or stats & probability subfields	WRPI score for level 2 standards in either the maths or stats & probability subfields	WRPI score for level 3 standards in either the maths or stats & probability subfields	Number of credits received with excellence at level 1	Number of credits received with excellence at level 2	Number of credits received with excellence at level 3	Cumulative score for level 1	Cumulative score for level 2	Cumulative score for level 3
Enrolled in a Tertiary qualification before 2016	1.00																		
Number of level 1 credits completed by age 18	0.38	1.00																	
Number of level 2 credits completed by age 18	0.38	0.66	1.00																
Number of level 3 credits completed by age 18	0.47	0.56	0.68	1.00															
expected percentile for level 1	0.46	0.70	0.54	0.56	1.00														
expected percentile for level 2	0.42	0.48	0.57	0.52	0.55	1.00													
expected percentile for level 3	0.40	0.40	0.46	0.59	0.45	0.63	1.00												
WRPI score for level 1 – all subjects	0.45	0.75	0.59	0.63	0.84	0.59	0.52	1.00											
WRPI score for level 2 – all subjects	0.46	0.60	0.68	0.69	0.73	0.66	0.57	0.86	1.00										
WRPI score for level 3 – all subjects	0.46	0.51	0.55	0.77	0.66	0.55	0.60	0.78	0.87	1.00									
WRPI score for level 1 standards in either the maths or stats & probability subfields	0.38	0.61	0.51	0.56	0.73	0.50	0.43	0.85	0.76	0.69	1.00								
WRPI score for level 2 standards in either the maths or stats & probability subfields	0.37	0.47	0.54	0.58	0.62	0.53	0.43	0.72	0.81	0.74	0.77	1.00							
WRPI score for level 3 standards in either the maths or stats & probability subfields	0.30	0.31	0.37	0.54	0.47	0.35	0.34	0.55	0.63	0.75	0.61	0.74	1.00						
Number of credits received with excellence at level 1	0.32	0.53	0.44	0.51	0.72	0.50	0.43	0.90	0.79	0.73	0.75	0.67	0.54	1.00					
Number of credits received with excellence at level 2	0.31	0.45	0.49	0.54	0.62	0.55	0.47	0.77	0.90	0.80	0.67	0.71	0.57	0.83	1.00				
Number of credits received with excellence at level 3	0.30	0.39	0.42	0.58	0.56	0.46	0.49	0.69	0.78	0.89	0.60	0.63	0.64	0.74	0.83	1.00			
Cumulative score for level 1	0.49	0.84	0.64	0.65	0.84	0.61	0.53	0.94	0.80	0.71	0.80	0.66	0.49	0.76	0.65	0.58	1.00		
Cumulative score for level 2	0.51	0.66	0.73	0.72	0.75	0.69	0.60	0.85	0.95	0.82	0.75	0.79	0.59	0.73	0.79	0.69	0.85	1.00	
Cumulative score for level 3	0.51	0.56	0.60	0.83	0.68	0.59	0.65	0.79	0.86	0.95	0.70	0.73	0.69	0.70	0.74	0.80	0.76	0.88	1.00

Source: Authors' calculations from Statistics New Zealand's Integrated Data Infrastructure (IDI).Note: N = 379,551. All observation counts have been randomly rounded as per Statistics New Zealand requirements.

Table 5: Correlations between WRPI, 2016 earnings, and other measures for all students born in 1988

	Earnings in 2016	Number of level 1 credits completed by age 18	Number of level 2 credits completed by age 18	Number of level 3 credits completed by age 18	expected percentile for level 1	expected percentile for level 2	expected percentile for level 3	WRPI score for level 1 – all subjects	WRPI score for level 2 – all subjects	WRPI score for level 3 – all subjects	WRPI score for level 1 standards in either the maths or stats & probability subfields	WRPI score for level 2 standards in either the maths or stats & probability subfields	WRPI score for level 3 standards in either the maths or stats & probability subfields	Number of credits received with excellence at level 1	Number of credits received with excellence at level 2	Number of credits received with excellence at level 3	Cumulative score for level 1	Cumulative score for level 2	Cumulative score for level 3
Earnings in 2016	1.00																		
Number of level 1 credits completed by age 18	0.16	1.00																	
Number of level 2 credits completed by age 18	0.17	0.63	1.00																
Number of level 3 credits completed by age 18	0.23	0.56	0.68	1.00															
expected percentile for level 1	0.21	0.69	0.55	0.59	1.00														
expected percentile for level 2	0.15	0.34	0.48	0.45	0.39	1.00													
expected percentile for level 3	0.14	0.33	0.42	0.54	0.38	0.62	1.00												
WRPI score for level 1 – all subjects	0.22	0.72	0.60	0.68	0.85	0.50	0.48	1.00											
WRPI score for level 2 – all subjects	0.21	0.53	0.66	0.69	0.71	0.57	0.51	0.85	1.00										
WRPI score for level 3 – all subjects	0.22	0.45	0.52	0.77	0.62	0.44	0.51	0.76	0.85	1.00									
WRPI score for level 1 standards in either the maths or stats & probability subfields	0.22	0.57	0.53	0.60	0.74	0.41	0.39	0.83	0.74	0.67	1.00								
WRPI score for level 2 standards in either the maths or stats & probability subfields	0.22	0.38	0.51	0.58	0.57	0.43	0.35	0.67	0.78	0.69	0.77	1.00							
WRPI score for level 3 standards in either the maths or stats & probability subfields	0.21	0.25	0.35	0.56	0.41	0.26	0.29	0.50	0.59	0.72	0.59	0.72	1.00						
Number of credits received with excellence at level 1	0.17	0.51	0.46	0.55	0.72	0.43	0.40	0.90	0.80	0.73	0.71	0.61	0.48	1.00					
Number of credits received with excellence at level 2	0.15	0.40	0.49	0.54	0.59	0.47	0.41	0.74	0.90	0.78	0.62	0.66	0.51	0.79	1.00				
Number of credits received with excellence at level 3	0.16	0.34	0.39	0.57	0.51	0.37	0.41	0.66	0.74	0.89	0.54	0.55	0.56	0.70	0.78	1.00			
Cumulative score for level 1	0.23	0.84	0.65	0.69	0.83	0.48	0.47	0.93	0.76	0.66	0.78	0.59	0.42	0.74	0.60	0.53	1.00		
Cumulative score for level 2	0.23	0.61	0.74	0.74	0.72	0.58	0.52	0.84	0.93	0.78	0.75	0.74	0.53	0.71	0.76	0.63	0.83	1.00	
Cumulative score for level 3	0.24	0.53	0.60	0.88	0.66	0.48	0.56	0.78	0.83	0.93	0.68	0.68	0.67	0.68	0.69	0.75	0.74	0.85	1.00

Source: Authors' calculations from Statistics New Zealand's Integrated Data Infrastructure (IDI).
Note: N = 10,833. All observation counts have been randomly rounded as per Statistics New Zealand requirements.

"With a smaller penalty for attempting more challenging courses, WRPI provides schools with weaker incentives to shunt students into less rigorous courses. The WRPI might thus be more suitable in exercises benchmarking relative school performance"

WRPI correlates most strongly with Cumulative Score, followed by the number of credits received with *excellence*, Expected Percentile, and number of credits completed.

Performance on WRPI generally correlates more strongly with tertiary enrolment than the number of credits received, or the number of credits received with *excellence*. The broad WRPI measure of overall NCEA attainment correlates more strongly with tertiary enrolment than the narrower measure restricted to maths, statistics and probability standards. The Cumulative Score at each level correlates with tertiary enrolment more strongly than WRPI.

Correlations using tertiary completion rather than tertiary enrolment for all students born between 1988 and 1993 showed broadly similar patterns.

We also tested correlations between student attainment measures and 2016 earnings for students born in 1988 (see Table 5). In 2016, those students would be about 28 years old and no longer in tertiary study.

WRPI correlates more strongly with later earnings than any attainment measure other than Cumulative Score despite minor differences and weak correlations overall.

Overall, WRPI generally correlates with tertiary enrolment, tertiary completion, and later earnings at least as well as other attainment measures. With a smaller penalty for attempting more challenging courses, WRPI provides schools with weaker incentives to shunt students into less rigorous courses. The WRPI might thus be more suitable in exercises benchmarking relative school performance.

We then ran regression analysis to determine the association between improved WRPI performance and tertiary enrolment, tertiary completion, and later life earnings, correcting for other characteristics available in IDI.

We tested how well WRPI could explain earnings in 2016 for students born in 1988. We corrected for gender, ethnicity and achieved tertiary qualifications using NCEA Level 1 scores as the WPRI score (Table 6). Future work could test the effects of the scores at different levels of study, and compare WRPI with other measures like the Cumulative Score and Expected Percentile.

Students born in 1988 would have completed NCEA during its early years; future work should check the consistency of results with cohorts completing NCEA after it had matured. The cohort born in 1992 would have been exposed to a more mature version of NCEA; they turn 28 in 2020.

Table 6: Correlates of earnings

Variable	Coefficient	SE	t-statistic	95% C.I.
WRPI score for level 1 – all subjects	48.1	5.8	8.3	37 – 60
Male	7,289.0	535.0	13.6	6,240 – 8,338
Ethnicity (omitted category – European)				
Maori	-2,537.5	697.3	-3.6	-3,904 – -1,171
Pasifika	-2,982.0	776.2	-3.8	-4,503 – -1,460
Asian	-6,142.6	748.6	-8.2	-7,610 – -4,675
MELAA*	-8,435.4	1,922.7	-4.4	-12,204 – -4,667
Other ethnicity	-8,746.6	2,586.6	-3.4	-13,817 – -3,676
Tertiary qualification field				
Natural and Physical Sciences	1,438.4	1,244.3	1.2	-1,001 – 3,877
Engineering and related technologies	10,930.1	1,654.5	6.6	7,687 – 14,173
Information Technology	8,003.8	1,882.3	4.3	4,314 – 11,693
Architecture and Building	3,237.4	2,293.0	1.4	-1,257 – 7,732
Agriculture, Environmental and Related Studies	6,229.3	2,895.9	2.2	553 – 11,906
Health	14,360.3	1,180.9	12.2	12,045 – 16,675
Education	10,307.6	888.7	11.6	8,566 – 12,050
Management and Commerce	7,458.9	960.9	7.8	5,576 – 9,342
Society & Culture	1,242.6	762.4	1.6	-252 – 2,737
Creative Arts	-3,301.2	880.1	-3.8	-5,026 – -1,576
Food, Hospitality and Personal Services	-4,180.5	1,782.5	-2.4	-7674 – -686
Mixed Field Programmes	100.4	3,861.4	0.0	-7468 – 7669
Constant	25,049.4	539.0	46.5	23993 – 26106

Source: Authors' calculations from Statistics New Zealand's Integrated Data Infrastructure (IDI). OLS regression. Dependent variable: Earnings in 2016. N=15018. F=44.96. R-squared = 0.0571.

Note: All observation counts have been randomly rounded as per Statistics New Zealand requirements. *Middle Eastern/Latin American/African.

"A standard deviation increase in WRPI, moving from the median score to the score at the 68th percentile, is associated with a 134% increase in the chances of enrolment"

A unit increase in WRPI increases earnings by $48.12. The mean WRPI score for students born in 1988 was 66.4 with a standard deviation of 59.1. This means a standard deviation increase in WRPI, or moving from the median score to the 68th percentile, is associated with a $2,844 increase in earnings at age 28.

We could not test the effects on earnings later in life as NCEA data is still too recent. But initial results may understate the differences in later earnings. Many 28-year-olds with higher scores at secondary school will have completed tertiary study and started gaining work experience, while many with lower scores would not have completed tertiary study, would have gained on-the-job experience, and may expect flatter future earnings growth. Future work should explore the return to university study, and to polytechnic study, adjusting for underlying student ability using measures like WRPI.

Correcting for gender and ethnicity, a unit increase in WRPI at NCEA Level 1 is associated with a 2.3% increase in the likelihood of enrolling in tertiary study. A standard deviation increase in WRPI, moving from the median score to the score at the 68th percentile, is associated with a 134% increase in the chances of enrolment (see Table 7).

Table 7: Correlates of tertiary enrolment

	Odds Ratio	SE	Z-statistic	95% Confidence Interval
WRPI score for level 1 - all subjects	1.023	0.000	204.380	1.023 - 1.024
Male	0.745	0.006	-38.670	0.734 - 0.756
Ethnicity (European omitted)				
Maori	0.608	0.006	-48.390	0.596 - 0.620
Pasifika	0.868	0.011	-11.070	0.847 - 0.890
Asian	1.838	0.025	45.180	1.790 - 1.888
MELAA	1.718	0.048	19.390	1.626 - 1.814
Other ethnicity	1.345	0.058	6.900	1.236 - 1.462
Cohort (year of birth)	0.767	0.001	-168.100	0.765 - 0.770
Constant	0.908	0.009	-9.290	0.890 - 0.927

Logistic regression. Dependent variable: Enrolled in tertiary education before 2016. All students born 1988–1997, n=401,670. Pseudo R2 = 0.2538. All observation counts have been randomly rounded as per Statistics New Zealand requirements.

"A unit increase in the WRPI score is associated with just under 1% increase in the likelihood of tertiary completion; a one-standard-deviation increase is associated with a 46% increase in the likelihood of tertiary programme completion"

Among those who enrol in tertiary study, those with higher WRPI scores are more likely to complete their programme. There, a unit increase in the WRPI score is associated with just under 1% increase in the likelihood of tertiary completion; a one-standard-deviation increase is associated with a 46% increase in the likelihood of tertiary programme completion (see Table 8).

Table 8: Correlates of tertiary completion, among those who enrol in tertiary study

	Odds Ratio	SE	Z-statistic	95% Confidence Interval
WRPI score for level 1 - all subjects	1.007	0.000	65.110	1.007 - 1.008
Male	0.729	0.009	-26.270	0.712 - 0.746
Ethnicity (European omitted)				
Maori	0.588	0.012	-27.110	0.566 - 0.611
Pasifika	0.485	0.012	-29.340	0.463 - 0.509
Asian	1.088	0.018	5.020	1.053 - 1.124
MELAA	0.876	0.036	-3.200	0.808 - 0.950
Other ethnicity	1.106	0.073	1.520	0.971 - 1.259
Cohort (year of birth)	0.735	0.003	-89.740	0.730 - 0.740
Constant	3.213	0.057	66.180	3.103 - 3.325

Logistic regression. Dependent variable: Completed a tertiary degree before 2016. All students born 1988–1994 who enrolled in tertiary study, n=274,755. Pseudo R2 = 0.2232. Note that observation numbers have been randomly rounded in keeping with Statistics New Zealand requirements.

The results in Tables 7 and 8 also show that, controlling for NCEA achievement, men are less likely than women to enrol in tertiary study, and are less likely to complete their qualification if they do enrol. Maori and Pasifika students are less likely to enrol in tertiary study while Asian and MELAA students are more likely to enrol. Conditional on enrolment, and correcting for NCEA achievement, Maori, Pasifika and MELAA students are less likely to complete their studies than European and Asian students.

As always, correlation does not imply causation, and the empirical work here only provides correlation. Students could have different WRPI scores because of different underlying abilities, quality of instruction, or assistance at home. Although higher WRPI scores are associated with higher earnings, and higher likelihood of joining and completing tertiary study, initiatives to boost NCEA attainment as measured by WRPI might not improve any of those outcomes. More sophisticated empirical methods, or trials of proposed initiatives, are needed to establish causal effects.

Nevertheless, WRPI's performance is encouraging.

Chapter 1 made the analytical case for an improved measure of student achievement in NCEA to mitigate the perverse incentives NCEA completion rates introduce: a school can improve its NCEA completion

"WRPI avoids the problems with existing measures while providing a student achievement score that not only correlates well with existing measures but also tracks future tertiary participation, tertiary qualification completion, and future earnings"

rate while worsening educational outcomes for its students by shunting students into easier classes.

Better information on transcripts would help employers and universities judge students' relative ability, but using the Expected Percentile could provide perverse incentives too. Students may avoid more challenging courses that worsen their Expected Percentile scores.

WRPI avoids the problems with existing measures while providing a student achievement score that not only correlates well with existing measures but also tracks future tertiary participation, tertiary qualification completion, and future earnings.

It could thus provide a better basis for comparing outcomes across schools, if adjusted appropriately for student background characteristics.

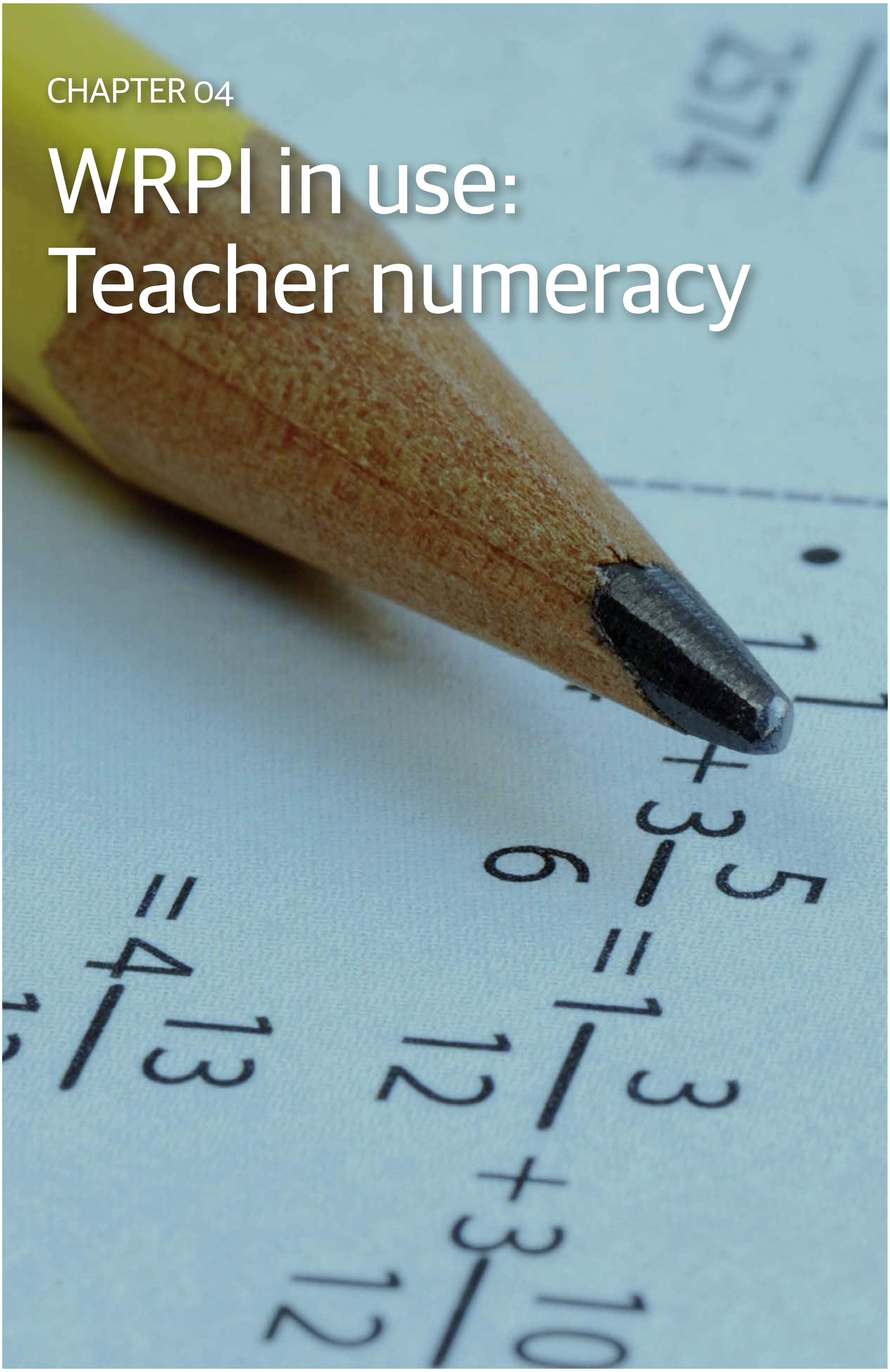
It could also form the basis for a more sophisticated investigation of the returns to various forms of tertiary education. It is trivial but true that engineering degrees lead to higher earnings than degrees in food service. Very different students pursue different degrees. More relevant is how much each degree improves lifetime earnings for students relative to other paths that they, or other students like them, chose.

WRPI provides an improved measure of student baseline attainment. Further work could augment it to account for differences in difficulty across different NCEA subjects.

Chapter 4 provides a case study use of WRPI.

CHAPTER 04

WRPI in use: Teacher numeracy



"The aim of the present report is not to discuss teacher numeracy, but it does present an opportunity to showcase how WRPI can be used to start answering pertinent policy questions"

Concerns about student performance in maths were documented in depth in a 2015 Initiative report, *Un(ac)countable: Why Millions on Maths Returned Little* (see Appendix A for the report summary).[32] Sector interviews indicated low maths ability among primary school teachers could have contributed to worsening outcomes in maths among students.[33] International benchmarking tests since *Un(ac)countable* show the concerns remain relevant.

The aim of the present report is not to discuss teacher numeracy, but it does present an opportunity to showcase how WRPI can be used to start answering pertinent policy questions.

We use WRPI to check whether the numeracy of people completing teaching degrees has declined. If it has, it could explain declining student performance. But we find the numeracy of graduates from teaching programmes has been increasing rather than decreasing, at least as indicated by those graduates' performance in numeracy NCEA subjects.

A tale of numeracy mediocrity

The numeracy of students from around 60 education systems is assessed every three to four years in international benchmarking tests.

Since 1994, the Trends in International Mathematics and Science Study (TIMSS) has assessed maths and science achievement of Years 5 and 9 (typically 10- and 14-year-olds). Questions in TIMSS are curricula aligned and designed to reflect the skills and knowledge taught in participating countries.[34]

Since 2000, the Programme for International Student Assessment (PISA) has assessed how thousands of 15-year-olds in OECD countries apply reading, science and maths skills and knowledge to real-life problems.[35]

Poor international rankings

New Zealand remains either below the international centre-point or has declined (Figure 5). TIMSS results show we have barely improved on the global stage in the last two decades. Year 5 students began dismally in 1994; after 20 years, while average scores have increased they remain below the international centre-point. Year 9 students' inaugural scores were at the international centre-point but have since dipped slightly (but statistically insignificantly). PISA performance for 15-year-olds has been declining in recent years.

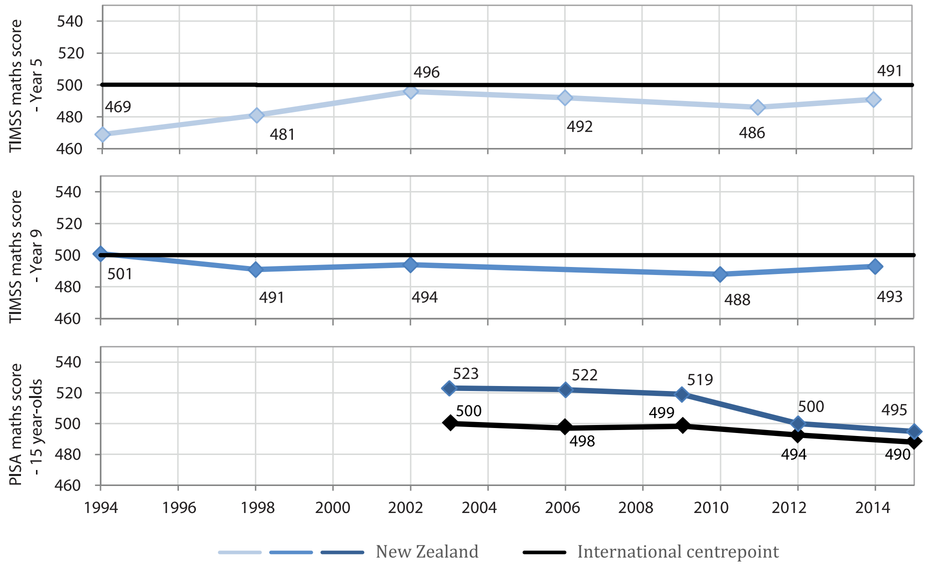
32 Rose Patterson, "Un(ac)countable: Why Millions on Maths Returned Little," op. cit.

33 Ibid. 29.

34 Education Counts, "TIMSS (Trends in International Mathematics and Science Study), 2014/15" (Wellington: Ministry of Education, various).

35 Education Counts, "PISA (Programme for International Student Assessment) 2015" (Wellington: Ministry of Education, various).

Figure 5: Trends in maths achievement for TIMSS and PISA (Mean scores)



Source: Adapted from Comparative Education Research Unit, "Mathematics Achievement: What We Know from New Zealand's Participation in TIMSS 2014/15 and PISA 2015," (Wellington: Ministry of Education, 2017), Exhibit 1; and Steve May, with Jonathan Flockton, and Sarah Kirkham. "PISA 2015 New Zealand summary report". (Wellington: Ministry of Education, 2016)
Note from publication: A score of 500 is not the same across the year levels. However, within each year, a score of 500 is the same across assessment years.

The poor performance across the two tests continues despite myriad programmes tackling maths performance such as the Numeracy Development Project (the Numeracy Project), which was the focus of *Un(ac)countable*. The $70 million professional development initiative introduced in 2001 a new way of teaching primary school maths and shifted the learning approach from instrumental to relational.[36] However, there is no evidence of the new approach lifting maths achievement. In fact, Year 5 students performed much lower in TIMSS 2010–11 than their counterparts in 2002–03, and Year 9 maths declined to below 1994 levels.[37] In the most recent test, Years 5 and 9 had the lowest mean score compared to Australia, England, United States, and Canada. Countries opt in and out of the tests and among the 13 countries that remained since 1994, New Zealand ranked the lowest in 2014.

Even more concerning is the proportion of students unable to solve basic maths questions. For the achievement range between advanced and poor performers, see Figure 6.

In TIMSS, students at the 'low' benchmark had some knowledge of whole numbers and basic graphs, while those below the 'low' benchmark could not perform the simple tasks designed for their grade and age cohort. In 2014, about 40% of Years 5 and 9 students were at or below the 'low' achievement benchmark (with about 15% below 'low' in each year group).

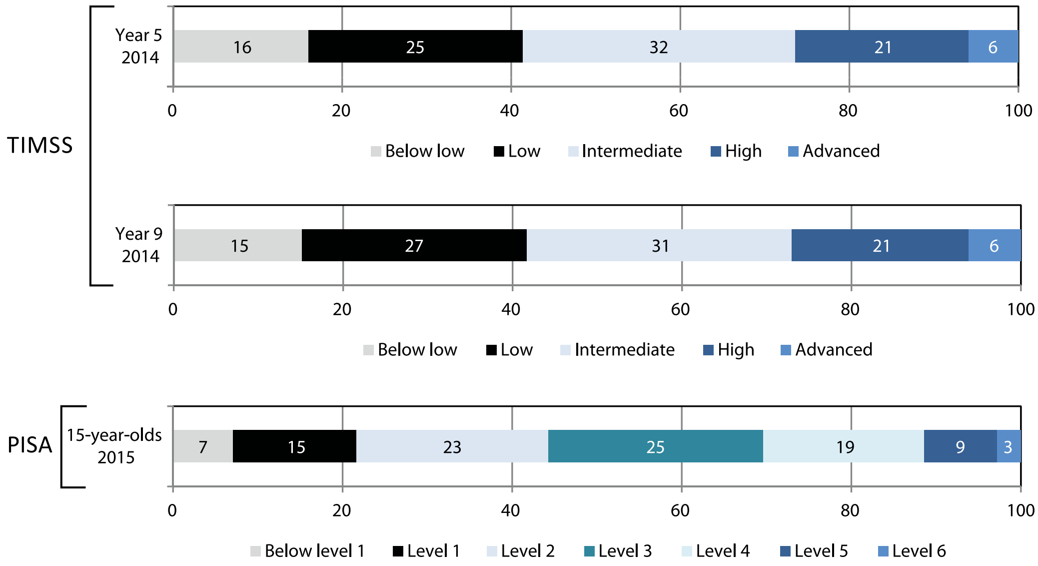
36 Rose Patterson, "Un(ac)countable: Why Millions on Maths Returned Little," op. cit. The report concluded that although both approaches are important, "the pendulum swung too far" towards relational instruction, and the balance was compromised.

37 Comparative Education Research Unit, "Mathematics Achievement: What We Know from New Zealand's Participation in TIMSS 2014/15 and PISA 2015" (Wellington: Ministry of Education, 2017), Exhibit 2.

To a degree, New Zealand's performance is worsening, with more Year 9 students performing at the lowest competency – from 11% in 1994 to 15% in 2014. This is the case even when the proportions of the younger Year 5 cohort reaching high and advanced benchmarks increased between 2010 and 2014.

The trends are as worrying for 15-year-olds in PISA. More students are achieving at the lowest band and fewer at the higher bands. In 2003, 15% students were below Level 2 in 2003 compared to 22% in 2015. At the top end, 21% students were at Level 5 or above in 2003 compared to 11% in 2015.[38]

Figure 6: Proportion of low and high achievers in maths (TIMSS and PISA)



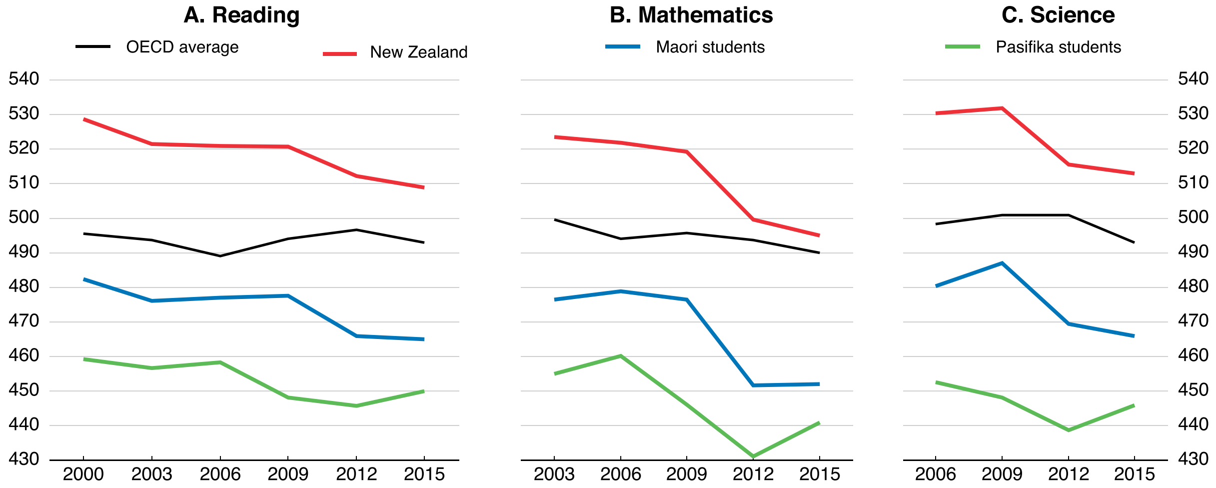
Source: Comparative Education Research Unit, "Mathematics Achievement: What we know from New Zealand's participation in TIMSS 2014/15 and PISA 2015," (Wellington: Ministry of Education, 2017), Exhibit 2. Note from publication: Benchmarks and proficiency levels are not comparable across year levels, but are indicative of the range of achievement within New Zealand.

In addition to the absolute weakness in maths, students have a relatively poor understanding of maths compared to science and reading (Figure 7).

New Zealand schools are clearly not equipping students with the same level of knowledge as in previous decades in international comparisons.

38 Performance at Level 2 in PISA is regarded as the baseline at "which students begin to demonstrate the competencies that will enable them to participate actively in maths-related life situations". Ibid.

Figure 7: Falling PISA scores in maths versus reading and science (2000–15)

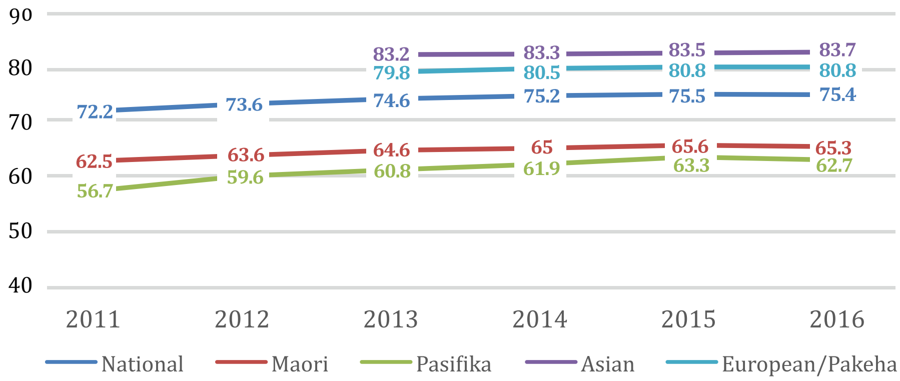


Source: OECD, PISA Results, various years; S. May, J. Flockton and S. Kirkham (2016), PISA 2015 – New Zealand Summary Report, Ministry of Education *cited in* OECD, "2017 OECD Economic Survey of New Zealand: Boosting productivity and adapting to the changing labour market," Presentation. OECD Publishing: Paris, 2017),

Weak national results

National Standards for primary and intermediate schools were introduced in 2010 to inform teachers and parents about student progress against expectations in reading, writing and maths.[39] A few more primary students either reached or exceeded the National Standards in 2016 than in 2011 (see Figure 8).

Figure 8: Percentage of students at or above the maths National Standards by ethnicity (2011–16)



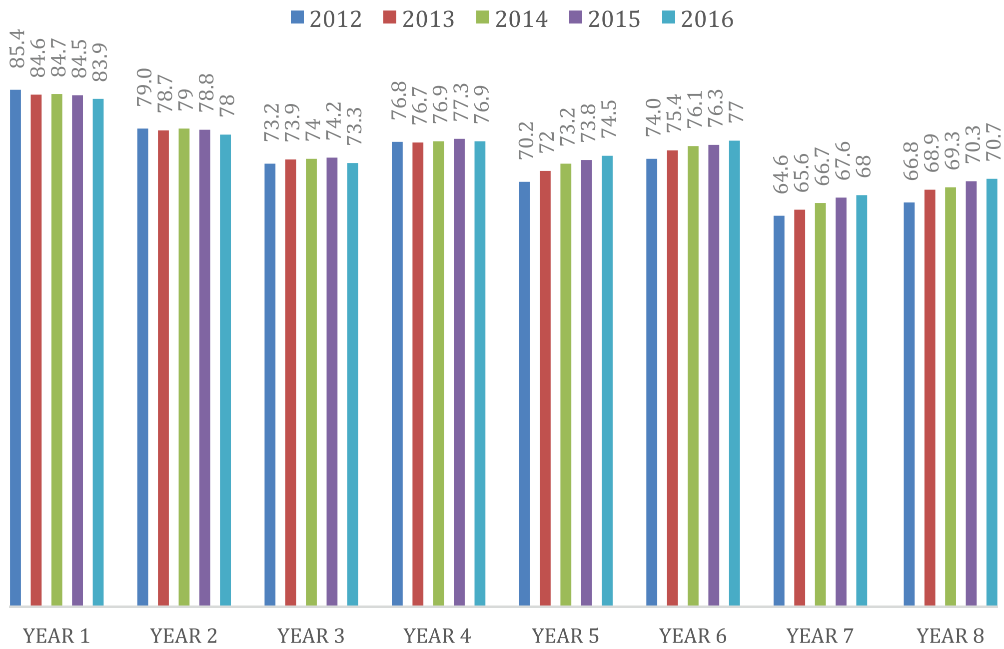
Source: Data for 2011–12 from Education Counts, "Public Achievement Information (PAI): New Zealand Education," various infographics, Website. Data for 2013–16 from Education Counts, "National Standards," Website.
Note: The Ministry of Education does not hold 2011 or 2012 National Standards data for Asian or European/Pākeha ethnic group students.

[39] Ministry of Education, "National Standards," Website. The Labour-led Government scrapped National Standards starting from 2018.

However, breaking down achievement by year shows a different story. Fewer students were assessed as meeting or exceeding the Standards as they moved through their schooling (see Figure 9).[40] In 2012, 85.4% students were assessed as 'at or above' the Standards by the end of their first year of school. The figure drops to 78.6% for Year 2 students in 2013 and 74% for Year 3 students in 2014. If we take the 2012 Year 4 cohort to be the 2016 Year 8 cohort, the story does not improve at the end of intermediate. In 2012, 76.8% students met or exceeded the Standards by the end of Year 4, but this figure was 70.7% by the end of Year 8.[41]

A 2013 national study of maths understanding found even worse results, with only 41% of Year 8 students at the expected year level, despite the majority achieving well just four years earlier.[42]

Figure 9: Proportion of students achieving at or above the National Standard for maths, by year level (2012–16)



Source: Education Counts, "Mathematics/Pangarau: primary schooling" Website.

A plausible explanation

Plausible reasons abound for New Zealand's dismal performance in maths. A series of reports in the past decade have identified several related issues: changes under the Numeracy Project; teacher confidence and subject knowledge; student confidence and perceptions; and decisions on professional development and training.

40 Education Counts, "Mathematics/Pangarau: primary schooling," Website.
41 Some reviewers found the "at Standard" benchmark to be vague in meaning.
42 Ministry of Education, "National Monitoring Study of Student Achievement: Mathematics and Statistics 2013," NMSSA Report 4 (Wellington: New Zealand Government, 2015).

In June 2017, two reports called for improving the quality of maths teaching. The Education Council released a consultation document for reviewing the provision of initial teacher education. It noted anecdotes from the sector that some teachers do not have adequate numeracy and literacy skills to teach, but had limited evidence about the size and scale of the problem. The Education Council proposes to increase the numeracy requirements for entry into teaching programmes.[43]

Similar concerns and proposals were listed in the OECD's two-yearly economic report on New Zealand.[44] The OECD recommended New Zealand improve maths teaching and professional development; review the minimum numeracy requirements for school qualifications; and raise the bar for entry into initial teacher training.[45]

Observations from the limited research on the numeracy of those in teacher training are grim.

In 2012, Jenny Young-Loveridge, Brenda Bicknell, and Judith Mills assessed the numeracy competency of 319 students enrolled in primary teacher education (Pre-service).[46] The majority (248) were pursuing their teaching qualification via a three-year Bachelor of Teaching degree. The rest were in a one-year Graduate Diploma of Teaching.

Students were given nine numerical tasks, with two at Level 3 of *The New Zealand Curriculum* and seven at Level 4. The levels are usually acquired by the end of Year 6 (10- to 11-year-olds) and Year 8 (12- to 13-year-olds), respectively. The undergraduate students were given the questions during orientation while the graduate students were assessed in their first class.

Pre-service teachers showed a weak ability to solve some basic maths questions (Table 9). Overall, 17% of the students scored less than 50% and 27% of the undergraduate students without University Entrance were in this group.

[43] The Education Council, "Educating Teachers for Our Changing World," Discussion paper (Wellington: 2017).

[44] OECD, "2017 OECD Economic Survey of New Zealand," op. cit.

[45] While the aim of this report is not to discuss research on teacher quality broadly, interested readers could peruse literature empirically linking teacher quality with student achievement in that differences in teacher performance can explain a large portion of student learning outcomes in school and beyond. See, for example, Raj Chetty, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review 104:9* (2014), 2633–2679. Others have reviewed the declining academic aptitude of incoming teachers. See, for example, Andrew Leigh and Chris Ryan, "How and Why Has Teacher Quality Changed in Australia?" *Australian Economic Review* 41:2 (2008), 141–159.

[46] Jenny Young-Loveridge, Brenda Bicknell, and Judith Mills, "The Mathematical Content Knowledge and Attitudes of New Zealand Pre-Service Primary Teachers" (The University of Waikato) *Mathematics Teacher Education and Development* 14:2 (2012), 28–49.

Table 9: Pre-service teachers' maths performance (percentage answering correctly, 2012)

Question	Undergraduates			Graduates
	University Entrance (n:150)	No University Entrance* (n:98)	Total (n:248)	n: 71
Level 3				
Tama has 64 stickers. He uses 27 on the first day of school. How many does he have left?	95	96	95	96
John needs $403 to buy a stereo. He has saved $297. How much money does he still need?	86	85	86	90
Level 4				
Sue used 8.3 metres of red material and 2.57 metres of blue material to make costumes for the play. How much material did she use altogether?	91	84	88	92
Ana bought 4.3 metres of rope to make skipping ropes, but only used 2.89 metres. How much rope was left over?	66	55	62	76
If 18 packets each hold 24 felt pens, how many pens is that altogether?	65	62	64	69
If 56 plums are shared among 14 people, how many plums will each person get?	92	91	92	89
Tama and Karen buy two pizzas. Tama eats 3/4 of one pizza while Karen eats 7/8 of the other one. How much pizza do they eat altogether?	37	22	32	56
If Ben got 72 out of a possible total of 90 marks, what percent-age was that?	31	24	28	48
Jo spent $60 on stationery. She got one-third off the original price, because she was a teacher. What was the original price?	65	56	62	72

Source: Jenny Young-Loveridge, Brenda Bicknell, and Judith Mills, "The Mathematical Content Knowledge and Attitudes of New Zealand Pre-Service Primary Teachers" (The University of Waikato) Mathematics Teacher Education and Development 14:2 (2012), 28–49, Table 2. *56 had been given Special Admission, 34 had come from other Tertiary institutions, and 8 had Discretionary Entry.

"Using WRPI and linked administrative data in Statistics New Zealand database, IDI, we can attempt to answer this question and demonstrate how such information can help policymakers diagnose persistent public policy problems like school underachievement"

Given the questions were within the curriculum level the aspiring teachers would teach, these results are troubling. Could the numeracy of primary school teachers explain the worsening student achievement in maths?

Using WRPI and linked administrative data in Statistics New Zealand database, IDI, we can attempt to answer this question and demonstrate how such information can help policymakers diagnose persistent public policy problems such as school underachievement.

How numerate are teachers?

Declining teacher numeracy is one potential explanation for declining student performance on international benchmarking tests. We consequently looked at whether there were improvements or declines in the numeracy of incoming teachers since NCEA. Using WRPI with NCEA data for anyone born between 1988 and 1994,[47] we observed:

- differences in numeracy and maths scores for students who enrol in education versus other majors;
- differences in numeracy and maths scores for students who finish teaching qualifications compared to the general population; and
- time trends for the full sample.

Changes in university funding under the Key National Government prioritised STEM disciplines, so it is plausible more of the numerate students were pulled into STEM. Fewer of the most numerate students would remain for other disciplines, including education. This could matter if newer pedagogies required more numerate primary teachers.

To check whether the numeracy of teachers has been changing required first identifying teachers in the IDI and then determining their WRPI maths score from NCEA results.

Identifying teachers in IDI

Although the IDI has vast potential for social research and other policy areas, it has significant hurdles in identifying teachers, specifically primary school teachers. The key hurdle is that employment information on individuals is not readily available to all researchers.

Though it is possible to identify an individual's employers through the Inland Revenue tax data, non-government researchers are restricted from seeing information that identifies businesses or organisations. Hence, a private organisation cannot access this information unless contracted by a government department.[48]

Other official data identifying individual employment information are of limited use for our purposes. Data are collected from the Census or a sample (e.g. the Household Economic Survey, Household Labour Force Survey, New Zealand Income Survey, or the Survey of Family Income and Employment). However, the number of teachers who can be linked to education data in these surveys is far too low to draw reliable conclusions. The census too is conducted infrequently – the last one was held in 2013.[49]

Since the IDI captures qualification enrolment and completion data for all tertiary students since 2007, we focus on those on the path to a teaching job: those who have enrolled in and/or completed a tertiary qualification enabling them to teach.

However, while identifying students who have majored in education is an easy task, it is not easy from IDI to determine who was enrolled

47 The birth cohort was selected to capture the largest sample with the most reliable NCEA results. Those born in 1988 would have taken NCEA when it was introduced between 2002 and 2004 while those born in 1994 would be 22 in 2016.

48 Access provisions have since changed to allow tertiary academic researchers access to LBD and full business tax data on a case-by-case basis. Non-government, non-university-affiliated researchers would not have access unless under contract to a government department.

49 The Census is typically carried out every five years; however, the 2011 Christchurch Earthquake postponed the one scheduled for that year and was conducted in 2013.

"Contrary to our hypothesis, we found no evidence of declining maths ability among aspiring primary teachers compared to the two other cohorts"

in a school teaching qualification. An education major could include qualifications to teach tertiary, adult, early childhood, second-language, or special needs students or general-purpose education study.

Teach NZ provides information on the qualification pathways to become teachers.[50] However, this list is not exhaustive, cannot be fully linked with qualifications in the IDI, and does not necessarily cover qualifications offered in the past.

We used the Ministry's method to identify ITE students and graduates.[51] This involved using the New Zealand Classification of Education (NZSCED) codes to identify classified primary and secondary teaching qualifications. Where ITE classifications are not clear, primary, secondary or other qualification is determined by whichever sector the student has undertaken most of their courses in.

Using this process, out of 5,586 teaching graduates we identified 741 'ambiguous' students who could not be categorised as either primary or secondary. Aspiring teachers born between 1988 and 1994 were identified and linked to their NCEA results (until the end of 2016).

Aspiring teachers' numeracy

Although Levels 1, 2 and 3 credits are typically sat in Years 11, 12 and 13 respectively, students can sit any NCEA credits any time during high school and/or after school. We looked at NCEA Level 1 results against 2016 regardless of when the credits were sat for two main reasons. First, school is compulsory until age 16, so most individuals in our cohort would have completed at least Level 1. Second, aspiring teachers who struggled with numeracy at the level equivalent to the first year of senior high school may likely struggle to teach primary school maths. Numeracy credits are a prerequisite to gain NCEA Level 1, and struggling at this level is the lowest bar available. (See Appendix B for NCEA Level 2 and 3 results.)

For our group of interest, we looked at the number of NCEA credits taken and passed. We constructed a maths performance score and an overall NCEA performance score for each individual – and produced a WRPI score for 3,384 aspiring primary and 1,461 secondary teachers, and for 85,095 other equivalent tertiary graduates.

Aspiring teachers look all right

Contrary to our hypothesis, we found no evidence of declining maths ability among aspiring primary teachers compared to the two other cohorts.

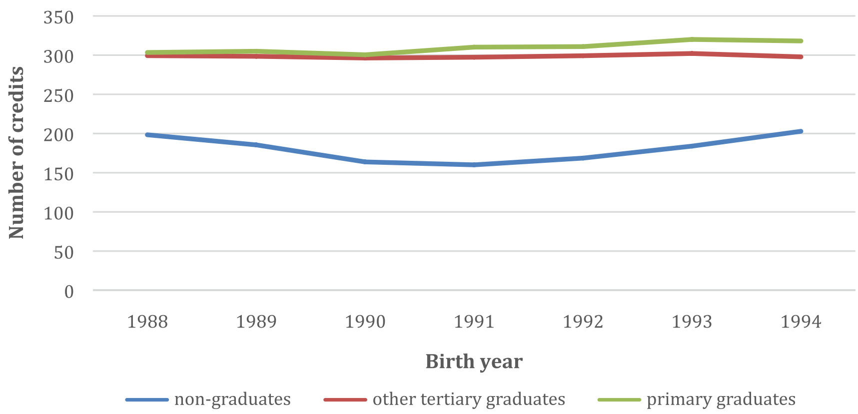
The following figures compare graduates of teaching programmes aimed at primary school teachers with graduates of other tertiary programmes and with those who do not go on to graduate from tertiary study. This last category would include both those who do not enrol in tertiary study and those who enrol but who do not complete a degree.

Aspiring primary teachers have consistently taken slightly more NCEA credits than other tertiary graduates and substantially more than non-graduates (see Figure 10).

50 Teach NZ, "Qualifications information for 2017," Website.

51 Education Counts, "Initial Teacher Education," Website.

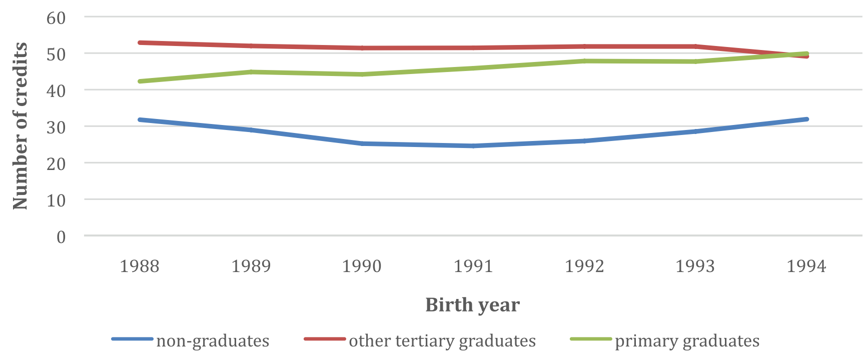
Figure 10: Total NCEA credits taken



Although enrolling in many standards may signal motivation, it does not reveal anything about consequent performance. It is more useful to look at the percentage of credits passed. In NCEA Level 1, primary graduates passed 87% of all credits taken compared to 90% for secondary graduates and 87% for equivalent tertiary graduates. The pass rate was 71% for the non-graduates (see Appendix B for results tables).

There is no indication that aspiring teachers have on average been enrolling in fewer maths credits than the comparison groups. In fact, younger primary graduates have been enrolling in slightly more maths credits than the equivalent tertiary graduates (see Figure 11). The 1988 other tertiary graduates cohort took on average about 10 more credits than the primary graduates. The gap has been closing since: the numbers trend upwards in 1994 for aspiring teachers and downwards for the other graduates cohort.

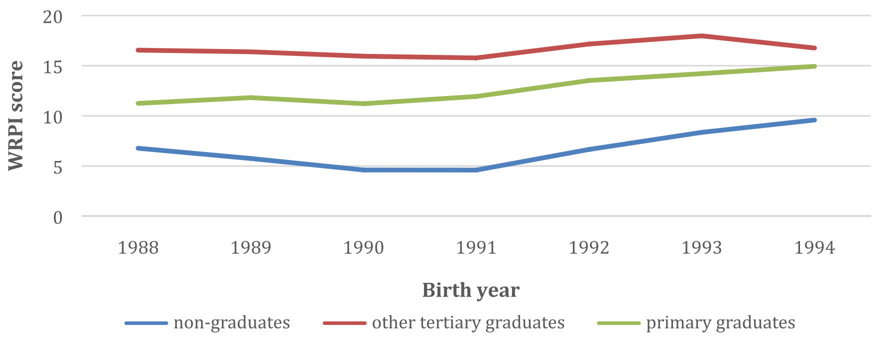
Figure 11: Total maths credits taken



Figures 12–13 illustrate our findings using WRPI to calculate maths and overall performance in NCEA Level 1 for the three comparison groups.

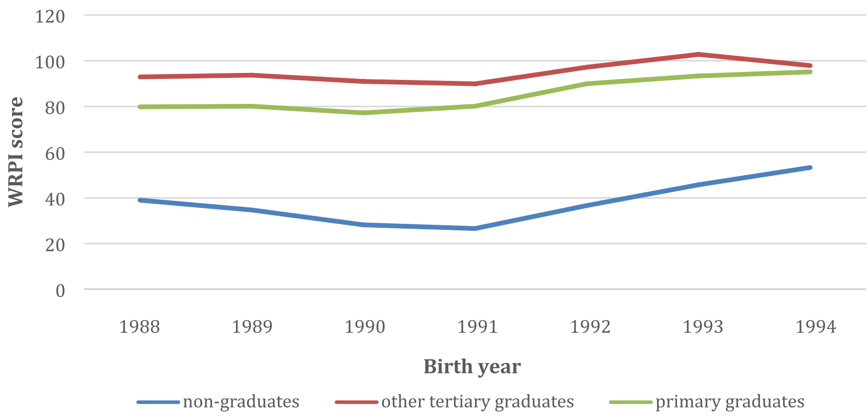
Apart from a slight dip for the 1990 and 1991 cohorts, the maths achievement of aspiring primary teachers, other graduates, and non-graduates has improved (see Figure 12).

Figure 12: NCEA Level 1 maths performance



A similar trend for aspiring teachers for overall NCEA ability is evident in Figure 13. The 10 or so WRPI score gap between primary and other graduates for the older (1988) cohort has narrowed to almost a 3 WRPI score-point difference for the 1994 cohort.

Figure 13: Overall NCEA Level 1 performance



New Zealand's maths performance is troubling but there has been no systematic evaluation of the numeracy of teachers or at least of those who graduate from teaching programmes. One reason may be that until a few years ago, data on high school results and post-school outcomes have not been as robust or been held in one accessible database – until Statistics New Zealand's IDI.

However, it remains challenging to quantify performance using NCEA outcomes due to its standards-based assessment structure, which can make students with vastly varied ability (and leaving school with the same qualification) look equally competent.

Using information about every individual in IDI and linking it to high school and university performance, the WRPI performance metric provides better information on relative competence than is being used in school achievement research. We applied WRPI to New Zealand's poor performance in maths to test changes in numeracy of those enrolling and completing initial teacher educational qualifications.

However, the evidence does not support the hypothesis that declining teacher numeracy explains the poor maths performance on international and national assessments.

“Based on NCEA Level 1 of people born between 1988 and 1994, we found no evidence of declining maths ability for teacher graduates, either absolute or relative to other tertiary graduates. These trends should be interpreted conservatively“

We find no evidence of declining NCEA Level 1 maths ability among graduates of teaching programmes, either in absolute terms or relative to other tertiary graduates. These trends should be interpreted conservatively as they assume the standard required for meeting NCEA credits has been maintained over the years of observation.[52] It is possible the upward trend could be a result of grade inflation or a general shift from external to internal assessments. While that is true, it does not explain convergence in numeracy between teaching and other graduates.

The stock of teachers greatly exceeds the annual flow of teachers: there are far more people employed in teaching than there are new teachers in any given year. Our WRPI case study looks only at the recent flow of teachers. So, if a change in pedagogy required higher standards in maths, it may be that many existing teachers were not up to the task. If so, continued improvement in the numeracy of incoming teaching graduates may mitigate the problem over time.

Although seven years of observation captures the largest window of reliable NCEA and IDI data, the period may not be long enough to make conclusive statements. We therefore urge the government to continue the evaluation as more data is collected.

This chapter shows the value of the IDI and using a better school performance metric to provide useful insights about a pertinent policy issue in New Zealand.

[52] We thank multiple reviewers for this point.

Conclusion

Accurately measuring student performance in NCEA is important. It helps students better understand their own strengths, parents to assess how their schools are doing, universities to admit students likely to succeed, and researchers to understand the causes and consequences of stronger student achievement.

However, the standards-based design of NCEA, New Zealand's main high school qualification, makes this a difficult task. Students sit different sets of standards, the difficulty of which can change dramatically within and between subjects, and in different years. Nevertheless, some research assesses the effects of education on later life outcomes simply by taking the highest attained qualification, such as NCEA Level 1, 2, or 3, as a measure of student performance.

We examined the limitations of available performance metrics and developed a new metric, the Weighted Relative Performance Index (WRPI), to measure student performance with NCEA data. Students with stronger performance on our WRPI measure prove more likely to enrol in tertiary study, to graduate, and enjoy higher earnings.

We used WRPI to see whether there have been changes in the numeracy ability of those enrolling in and completing initial teacher education qualifications. Low teacher numeracy has been suggested as a contributing factor to the decline in numerical ability of New Zealand school students as demonstrated in international rankings. Using the new metric, we see no evidence of declining maths ability among incoming teachers. However, we cannot rule out numeracy deficiencies in the existing stock of teachers as a potential explanation.

The measure could be used more broadly in other research on the causes and consequences of improved education performance. Further work could augment the measure to account more fully for differences in difficulty across courses.

Improving educational outcomes starts with measuring them. This report provides a step forward.

Appendix A

Executive summary of *Un(ac)countable): Why Millions On Maths Returned Little*[53]

Fifteen years ago, a wave rippled through New Zealand: The Ministry of Education introduced a new way of teaching maths in primary schools. The Numeracy Development Project (Numeracy Project) a centrally devised professional development (PD) programme for primary school maths teachers, was rolled out in 2001. The Numeracy Project changed the way maths is taught in New Zealand primary schools, putting more emphasis on teaching children multiple mental strategies for solving problems. It followed a series of smaller localised PD programmes in the mid- to late-1990s that showed signs of success. The Numeracy Project was intensive, with around 20 hours of PD for each primary school teacher in its first two years, and expensive, at a central cost of $70 million.

Maths performance has been in decline over the last 10 years, with losses in the basics

- Maths performance showed signs of improvement in the mid- to late-1990s, but has been in decline since then, although not back to early 1990s levels. There have been losses in the basics such as simple addition and multiplication, and children are no longer using vertical written methods for solving maths problems.
- It is a myth that children in the East Asian countries (that top the charts in maths) are just rote learning for the tests. Though they score highly on knowledge of basic facts, they are also better than New Zealand students at applying their knowledge to solve novel problems.

The Numeracy Project has put too much emphasis on multiple strategies, and not enough on the basics

- In tandem, curriculum changes over time show a move towards more 'relational' learning (discerning the connections between numbers and situations by mentally working out answers to maths problems, often using multiple strategies) over 'instrumental' learning (basic maths rules and processes using the traditional written form).
- Although the Ministry of Education maintains that both knowledge and strategies are important, children in New Zealand are spending more time explaining their answers in class and less time memorising facts, rules and procedures, compared to children in other countries, including the top performers.
- Relational learning is important, but so is gaining fluency in the basics and written methods, which frees up children's working memory to develop the deeper conceptual mathematical understanding the Numeracy Project intended.

53 Rose Patterson, "Un(ac)countable: Why Millions on Maths Returned Little," op. cit.

Many primary school teachers may not be maths proficient to teach the new methods

A 2010 study found that a third of new primary school teachers could not add two fractions (7/18 + 1/9). Yet today's emphasis on developing children's deeper conceptual understanding in maths may rely even more on teacher maths abilities.

Both teacher maths proficiency and maths teaching proficiency (knowledge of how to represent mathematical concepts in ways children can understand) are predictive of student achievement in maths. Yet there are no objective assessments of whether graduating teachers have the required level of proficiency. Teacher salaries relative to other professional occupations such as law, accounting, engineering and science have stayed stable over the past 15 years. This makes the explanation that declines in maths are due to declines in teacher maths abilities unlikely.

The cost of the Numeracy Project has not been worth the benefit

The Numeracy Project has returned little benefit at substantial cost. This report outlines the problems with imposing a centrally planned, nationwide approach to teaching maths on top of a self-managing education system. It asks who is accountable for results. Parents have been asking questions about the new methods of teaching maths, and schools too have begun to question the methods.

Recommendations for consideration

The Numeracy Project shows that centrally devised approaches to changing instruction are not appropriate, nor do they necessarily return the intended benefit. As such, The New Zealand Initiative proposes that individual schools should weigh up whether they have the right balance of instrumental and relational learning for maths, and make adjustments if necessary. This report also makes the following recommendations for consideration:

- The Investing in Educational Success (IES) policy presents an opportunity for teachers strong in maths to share their expertise with other teachers. Communities of Schools1 signing up for IES should consider how they can best share maths teaching knowledge.
- Schools in New Zealand adapt the national curriculum to each local context. The Ministry of Education should consider ways that the maths curriculums of successful schools can be shared with other schools serving similar student profiles.
- A certificate of maths teaching proficiency should be developed, based on a test of both maths ability and maths teaching ability (such tests, which validly predict maths teaching ability, have been devised overseas). This should not be mandated but be optional for teachers who want to gain their maths proficiency certification.

Appendix B

Results tables (Teacher numeracy)

	Percentage of credits passed									
			Enrolled				Completed			
	Full sample	Non-graduates	Primary teaching	Secondary teaching	General teaching	Equivalent tertiary	Primary teaching	Secondary teaching	General teaching	Equivalent tertiary
Observations	377,685	208,215	3,297	1,548	4,539	169,470	3,318	1,428	717	80,928
Numeracy NCEA 1 (% passed)	80%	74%	0.84	90%	84%	87%	86%	91%	84%	89%
	262,932	115,428	2,745	1,365	3,840	147,501	2,850	1,275	594	71,553
Numeracy NCEA 2	68%	58%	72%	77%	71%	76%	73%	79%	73%	78%
	155,172	51,000	1,557	924	2,394	104,172	1,695	864	324	49,398
Numeracy NCEA 3	72%	64%	73%	75%	73%	76%	74%	76%	76%	78%
	373,323	204,153	3,297	1,548	4,533	169,173	3,318	1,428	720	80,895
Core maths NCEA 1	80%	74%	83%	90%	84%	87%	86%	91%	84%	89%
	252,069	107,655	2,691	1,356	3,753	144,414	2,817	1,263	585	70,914
Core Maths NCEA 2	66%	57%	67%	73%	66%	73%	68%	74%	67%	74%
	126,744	33,963	1,398	882	2,103	92,781	1,650	849	312	48,555
Core Maths NCEA 3	75%	67%	77%	77%	76%	77%	78%	77%	82%	79%
	374,088	204,840	3,297	1,548	4,536	169,248	3,318	1,428	720	80,919
Maths & Statistics NCEA 1	80%	74%	84%	90%	84%	87%	86%	91%	84%	89%
	257,631	111,564	2,733	1,365	3,804	146,067	2,850	1,275	594	71,535
Maths & Statistics NCEA 2	68%	58%	71%	77%	71%	76%	73%	79%	73%	78%
	141,579	41,298	1,491	912	2,268	100,278	1,698	861	324	49,389
Maths & Statistics NCEA 3	72%	63%	72%	75%	72%	75%	74%	76%	76%	78%
	378,255	208,305	3,297	1,551	4,551	169,950	3,321	1,431	723	81,258
Maths, Stats & Science NCEA 1	74%	68%	78%	86%	78%	82%	81%	87%	78%	84%
	277,296	123,657	2,922	1,455	4,044	153,642	3,036	1,359	645	75,075
Maths, Stats & Science NCEA 2	64%	54%	67%	73%	67%	72%	69%	75%	69%	74%
	163,671	50,238	1,893	1,125	2,814	113,436	2,133	1,062	414	56,061
Maths, Stats & Science NCEA 3	66%	56%	65%	69%	65%	70%	67%	70%	67%	72%
	393,756	219,513	3,336	1,572	4,596	174,246	3,351	1,452	735	83,580
All NCEA Level 1	77%	71%	84%	89%	84%	85%	87%	90%	84%	87%
	380,904	207,543	3,339	1,572	4,593	173,358	3,354	1,446	738	83,430
All NCEA Level 2	77%	74%	80%	83%	79%	80%	81%	84%	80%	82%
	327,981	164,112	3,252	1,554	4,452	163,869	3,288	1,431	711	79,350
All NCEA Level 3	76%	76%	76%	79%	76%	77%	77%	80%	77%	78%

Bibliography

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review 104:9* (2014), 2633–2679.

Comparative Education Research Unit, "Mathematics Achievement: What We Know from New Zealand's Participation in TIMSS 2014/15 and PISA 2015," (Wellington: Ministry of Education, 2017).

Crichton, Sarah, Robert Templeton, and Sarah Tumen. "Using Integrated Administrative Data to Understand Children at Risk of Poor Outcomes as Young Adults," Analytical Paper 15/01 (Wellington: Treasury, 2015).

Education Counts. "Initial Teacher Education," Website (Wellington: New Zealand Government).

—. "Mathematics/Pangarau: primary schooling," Website (Wellington: New Zealand Government).

—. "New Zealand Schools: Ngā Kura o Aotearoa," Website (Wellington: New Zealand Government, 2014–2016 reports).

—. "PISA (Programme for International Student Assessment) 2015" (Wellington: Ministry of Education, various).

—. "TIMSS (Trends in International Mathematics and Science Study), 2014/15" (Wellington: Ministry of Education, various).

Joyce, Peter. "System fatally flawed," *The Christchurch Press*, cited in Muriel Newman, "How good is our education system?" (New Zealand Centre for Political Research, 4 June 2007).

Leigh, Andrew and Chris Ryan. "How and Why Has Teacher Quality Changed in Australia?" *Australian Economic Review* 41:2 (2008), 141–159.

Lipson, Briar. "Spoiled by Choice: How NCEA hampers education, and what it needs to succeed" (Wellington: The New Zealand Initiative, 2018).

Lyons, Peter. "NCEA pass rates hard for parents to cope with," *The New Zealand Herald* (25 November 2015).

Maloney, Tim and Kamakshi Singh. "Using Validated Measures of High School Academic Achievement to Predict University Success," AUT School of Economics *Working Paper Series*, 2017/10 (AUT School of Economics, December 2017).

Meehan, Lisa, Gail Pacheco, and Zoe Pushon. "Explaining ethnic disparities in bachelor's degree participation: evidence from NZ" Studies in Higher Education (2018), DOI: 10.1080/03075079.2017.1419340

May, Steve, with Jonathan Flockton, and Sarah Kirkham. "PISA 2015 New Zealand summary report". (Wellington: Ministry of Education, 2016)

Ministry of Education. "National Monitoring Study of Student Achievement: Mathematics and Statistics 2013," NMSSA Report 4 (Wellington: New Zealand Government, 2015).

—. "Review of Education Funding Systems," Cabinet Paper (Wellington: New Zealand Government, 2016).

New Zealand Institute of Economic Research (NZIER). "Strengthening the Government's Open Investments" (Wellington: NZIER, 2015).

New Zealand Qualifications Authority (NZQA). "NCEA," Website.

—. "NCEA endorsements: Recognising high achievement with 'endorsements'," Website.

—. "NCEA levels and certificates," Website.

—. "Secondary Statistics Consolidated Files," Website.

—. "Standards," Website.

OECD. "2017 OECD Economic Survey of New Zealand: Boosting productivity and adapting to the changing labour market," Presentation (Paris: OECD Publishing, 2017).

Patterson, Rose. "Un(ac)countable: Why Millions on Maths Returned Little" (Wellington: The New Zealand Initiative, 2015).

Rodgers, Lisa. "NCEA is not a pass or fail system," *The New Zealand Herald* (26 November 2015).

Social Investment Unit. "Briefing to the Incoming Minister" (Wellington: Ministry of Social Development, 2016).

State Services Commission. "Better Public Services 2012–2017," Website.

Statistics New Zealand. "How researchers are using the IDI: Education and training," Website.

—. "Integrated Data Infrastructure," Website.

Teach NZ. "Qualifications information for 2017," Website.

The Education Council. "Educating Teachers for Our Changing World," Discussion paper (Wellington: 2017).

Udahemuka, Martine. "Amplifying Excellence" (Wellington: The New Zealand Initiative 2017).

Ussher, Scott. "Post-School Choices: How Well Does Academic Achievement Predict the Tertiary Education Choices of School Leavers?" (Wellington: Education Counts, 2008).

Varian, Hal. *Microeconomic Analysis*, 3rd ed. (Norton, 1992).

Young-Loveridge, Jenny, Brenda Bicknell, and Judith Mills. "The Mathematical Content Knowledge and Attitudes of New Zealand Pre-Service Primary Teachers" (The University of Waikato) *Mathematics Teacher Education and Development* 14:2 (2012), 28–49.

How do we know how effective our schools are? What do we know about how well our students achieve, and how good are our teachers?

These questions have bamboozled, puzzled and frustrated academics, researchers, politicians, economists, teachers, principals, parents and learners for a long time.

The authors of this report have set themselves the uphill task of critically examining one of these measures – NCEA qualification outcomes. This is entirely appropriate and timely as government moves to use 'big data' evidence, including NCEA outcomes, to guide major policy and resourcing decisions.

The authors' proposed solution is a "Weighted Relative Performance Index (WRPI)," which effectively creates an overall score for each student based on the particular NCEA standards they have been assessed against. This, they argue, would enable users of the data to more credibly judge a student's success, and thus the effectiveness of schools and teachers in helping students gain valid qualifications.

There is no doubt we need more robust measures of school and student effectiveness than we currently have. The authors of this report have courageously initiated what will be a fascinating and consequential debate.

Bali Haque
Former Deputy CEO of New Zealand Qualifications Authority (NZQA)
Wellington

$25.00

ISBN
978-0-9951009-0-9 (print)
978-0-9951009-1-6 (online)

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

The New Zealand Initiative
PO Box 10147
Wellington 6143