

IN FAIRNESS TO OUR SCHOOLS

Better measures for better outcomes

Joel Hernandez

Foreword by Sir Peter Gluckman



THE NEW ZEALAND INITIATIVE

© The New Zealand Initiative 2019

Published by

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

Views expressed are those of the author and do not necessarily reflect the views of The New Zealand Initiative, its staff, advisors, members, directors or officers.

ISBN

978-0-9951208-2-2 (print)
978-0-9951208-3-9 (online)

RR56

Designed by Angela Whitney,
www.angelawhitney.com

Printing arranged by True North New Zealand Ltd



Attribution 4.0 International (CC by 4.0)

DISCLAIMER

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand.

The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author(s), not Statistics NZ.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

IN FAIRNESS TO OUR SCHOOLS

Better measures for better outcomes

Joel Hernandez

Foreword by Sir Peter Gluckman

About the New Zealand Initiative

The New Zealand Initiative is an independent public policy think tank supported by chief executives of major New Zealand businesses. We believe in evidence-based policy and are committed to developing policies that work for all New Zealanders.

Our mission is to help build a better, stronger New Zealand. We are taking the initiative to promote a prosperous, free and fair society with a competitive, open and dynamic economy. We are developing and contributing bold ideas that will have a profound, positive and long-term impact.

ABOUT THE AUTHOR



Joel Hernandez is a Policy Analyst at The New Zealand Initiative. He joined the Initiative after completing his Master's in Economics at Victoria University, where his work focused on productivity, labour economics and game theory. Before this, he completed a Bachelor of Science from the University of Otago, majoring in Microbiology.

Joel is the author of the research note, *Tomorrow's Schools: Data and evidence*.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank all those who have generously contributed their time and expertise to this report. First, I thank my incredible colleagues at the Initiative. Dr Eric Crampton, Mangai Pitchai, Dr Oliver Hartwich, Linda Heerink, Simone White, Chelsy Blair, Dr Bryce Wilkinson, and Briar Lipson – all of them not only contributed countless hours towards this project but also mentored and guided me through my first report at the Initiative. Second, I thank my expert reviewers, Andrew Webber, Dr Jeremy Clark, Dr Michael Johnston, and Wilbur Townsend – their subject expertise has greatly improved the quality of this report. Finally, I thank the three secondary school principals who gave us permission to use their school data in our project.

The author is solely responsible for the views, findings and conclusions expressed in this report, and for any errors and omissions.

CONTENTS

Foreword	05
Key Findings	07
Policy Recommendations Summary	08
Executive Summary	09
Introduction	11
CHAPTER 1	15
Value-added models	15
The solution to school evaluation – Lessons from overseas	15
Overview of value-added models	16
Robustness (reliability) across models	17
Which value-added model should New Zealand use?	17
CHAPTER 2	19
School performance tool	19
The Initiative's school performance tool	19
How the Initiative's school performance tool works	20
Data from the Integrated Data Infrastructure	20
Technical details of the Initiative's school performance tool	21
Model components	21
CHAPTER 3	25
School performance tool results	25
Summary of decile results	25
Individual school performance results	26
Individual school performance across deciles	28
Future of our school performance tool	30
CHAPTER 4	32
Policy applications	32
Complement to ERO	32
Annual reports	32
Qualitative aspects and internal contribution	33
Learning from the top performers	34
Evaluation of any government education policy	34
Evaluation framework	35
Opening the black box of education	36

Conclusion	37
Policy Recommendations	37
Appendix	39
Future research notes	40
Endnotes	42
Bibliography	47

Figures

FIGURE 1: An overview of the data available in the IDI	13
FIGURE 2: Education black box	20
FIGURE 3: School performance tool	20
FIGURE 4: Unadjusted and adjusted average performance of secondary schools within each decile based on each student's NCEA level 1 WRPI score (2008-17)	25
FIGURE 5: Hypothetical value-added scores for three schools	26
FIGURE 6: NSW: Value-added scores for the entire school population	27
FIGURE 7: Unadjusted and adjusted performance of New Zealand secondary schools based on each student's NCEA level 1 WRPI score (2008-17)	28
FIGURE 8: Distribution of school performance: University Entrance (unadjusted)	29
FIGURE 9: Distribution of school performance: University Entrance (adjusted)	29
FIGURE 10: NSW: School performance relative to all other government schools	32
FIGURE 11: NSW: School performance over time	33

Tables

TABLE 1: Student NCEA outcomes variables	21
TABLE 2: Student background characteristics variables	23
TABLE 3: Parental background characteristics variables	23
TABLE 4: School type variables	24
TABLE 5: Relevant colours and categories for Figures 8 and 9	29
TABLE 6: Six key drivers of success from high value-add schools in NSW	35
TABLE 7: Weights applied to NCEA grades	39

Foreword



Every parent wants the best for their child. In an advanced and liberal democracy, it is reasonable to expect that the education system should be organised to help deliver this. Societies invest a lot in education, from preschool to tertiary, through both public and private means. There are considerable differences between countries, however, in how that investment is allocated. To be sure, there is much debate on the implications of these differences.

Much of what we think we know about the drivers of quality education in New Zealand, at least in the eyes of the public, is based more on anecdote or retrospective assessment than relatively objective and prospective information, which may account for the level of debate. Indeed, one reason for the contested nature of education policy is that different stakeholders will understandably emphasise different influences on education outcomes, i.e. whether these are endogenous to the school or part of the environment in which they are situated. Yet we know that individual educational outcomes are due an interaction between the circumstances into which children are born, the support they receive from family or whānau and the broader community, and the quality of both the formal and informal education they receive. Educational outcomes are also strongly affected by other factors such as the child's emotional and physical health. The challenge is how to make sense of all these factors and their relative importance. Increasingly, New Zealand is wanting to understand all these factors in an integrated way to do the best by our next generation.

Today, in many countries, there is a major transition underway in public policymaking that

can help provide such information. Because of advances in data science, the ability has emerged for governments to use detailed and linked data sets to explore the impact of public services in unprecedented ways. Data itself is not informative; it needs to be appropriately analysed and put into context by experts to be useful. The algorithms and structure of models used to look at the data need to be unbiased and transparent. It is important that they are available for other experts to explore and validate. In a sector such as education, this is particularly important. Further, conclusions must be validated and interpreted in the context of societal values.

In New Zealand, part of this effort must be to use the increasing amount of data that exists both within the education system and more broadly through the globally innovative Integrated Data Infrastructure. The report that follows is but one exploratory step forward in applying citizen-based analytics to understand more about the system as a whole, while appropriately acknowledging the limitations of the data and the proxy measures used both on input and output.

It is reassuring that the analysis points to there being high-quality schools and high-quality educational experience across the socioeconomic gradient of communities in which schools are based. Clearly, such analysis will have important implications for how we perceive all of our schools in New Zealand. The type of approach piloted here could lead to helpful benchmarking for quality improvement, provided it is done in an agreed and appropriate manner. Together with the ERO system, such analyses should help many schools progress.

However, and most importantly, the analysis also points to the continued and critical challenge

for New Zealand: how to make better progress on issues of intergenerational disadvantage to improve the circumstance in which all children in New Zealand grow and learn.

Sir Peter Gluckman

ONZ KNZM FRSNZ FMedSci FRS

University Distinguished Professor

Head, Centre for Science in Policy, Diplomacy and Society (SciPoDS)

The University of Auckland

Auckland

Key Findings

Research from our school performance tool indicates that:

- The differences in school performance typically seen in NCEA league tables largely reflect differences in the communities those schools serve, not large differences in school quality or effectiveness.
- Once adjusted for differences in family background, the large performance differences between deciles disappear; however, high-performing and underperforming schools exist across all deciles.
- In particular, when evaluated on University Entrance, 42 decile 1 and 2 schools outperform 75% of every other secondary school in the country; in contrast, 9 decile 9 and 10 schools are at the bottom 25% of all secondary schools in the country.
- Accounting for differences in family background, approximately 80% of schools perform almost identically when evaluated across a wide range of NCEA metrics.

Policy Recommendations Summary

1. **Issue annual reports.** Annual reports containing insights gained from our school performance tool should be provided by the Ministry of Education to every principal and school board of every secondary school in New Zealand. This will require demand from parents and for additional resources to be directed by the Minister of Education.
2. **Allow the identification of individual schools in the IDI.** This can be achieved by revisions to rule 5.14.2 of Statistics New Zealand's *Microdata Output Guide*; reinterpretation of the Memorandum of Understanding and *Statistics Act 1975*; or amendments to the Act.
3. **Further development of our school performance tool in the IDI by the Ministry of Education.** Any further research should be made open-source, identical to what we have done for all the coding for this project.
4. **Implementation of our tool as part of an evaluation framework in any new government education policy through the Ministry of Education.**
5. **ERO to investigate any differences in institutional practice between low-, middle- and high-performing schools as identified by our tool.** This would be in addition to further research comparing the Education Review Office's conclusions with the conclusions gained from our tool.
6. **Integration of Te Rito, Edsby and Novopay data into the IDI to allow more comprehensive evaluation of school performance in the future.**

Executive Summary

School evaluation is an essential component of the education system. Identifying high-performing and underperforming schools is vital for building an evidence base for what works to improve education outcomes for students. However, the task of evaluating schools objectively and without bias is extremely complicated.

The complex interactions between family socioeconomic background and academic performance result in measures of student academic achievement that are a mix of family, teacher and school effects. National Certificate of Education Achievement (NCEA) league tables – which only show absolute measures of student achievement – tell us as much about students and their parents as the quality of the school.

When 35% of year 12 students in a school receive an Excellence endorsement at NCEA level 2, is that the result of a highly effective school that gets its students to meet and exceed their potential, the result of higher socioeconomic students enrolled in that school, or a combination of both?

A solution many countries have adopted is value-added models of assessment and evaluation. Value-added models differ from other methods of assessment in that they use a mix of prior student achievement data and family socioeconomic background data to separate the contribution of family background from the contribution of the school. This allows the Ministry of Education – in countries such as the United States, the United Kingdom and Australia – to identify what proportion of student achievement can be attributed to the school and to the family.

Additionally, value-added models allow the Ministry to identify how effective one individual

school is compared with every other school in the country, and which characteristics or institutional practices are associated with effective schools. The exact value-added model implemented varies depending on the purpose, stakes, political climate, and data availability in each country.

New Zealand does not have a nationwide value-added model, but it desperately needs one. Fortunately, New Zealand has an opportunity to build and implement one through Statistics New Zealand's (SNZ) Integrated Data Infrastructure (IDI) – New Zealand's largest research database. Better yet, because of the world-leading data available in the IDI, The New Zealand Initiative's value-added model improves upon existing models significantly in both the breadth and depth of data it uses to identify and separate the contribution of family background.

Over the past year, the Initiative has done just that – build New Zealand's first contextualised value-added model – alternatively called our school performance tool.

Using linked administrative data from the Ministry of Education, New Zealand Police, Ministry of Social Development, Department of Corrections, Ministry for Children, Immigration New Zealand, Inland Revenue, and the 2013 Census, the Initiative's school performance tool is able to identify how much each secondary school contributes to its students after separating the contribution of each student's family socioeconomic background.

This has allowed us to fairly and objectively compare low- and high-decile schools for the first time in New Zealand.

The results from our school performance tool indicate that the differences in school performance typically seen in NCEA league tables largely reflect the differences in the communities those schools serve, not large differences in school quality or effectiveness.

Additionally, once adjusted for differences in family background, the large performance differences between deciles disappear; however, high-performing and underperforming schools still exist across all deciles. In particular, when evaluated on University Entrance, 42 decile 1 and 2 schools outperform 75% of every other secondary school in the country; in contrast, 9 decile 9 and 10 schools are at the bottom 25% of all secondary schools in the country.

Furthermore, accounting for differences in family background, approximately 80% of schools perform almost identically when evaluated across a wide range of NCEA metrics.

The purpose of the Initiative's school performance tool is to show what can and should be done with the world-leading data available

in the IDI. The insights gained from our tool should be provided to all schools, principals and boards of trustees by the Ministry of Education.

If used by the Ministry, our tool could provide annual reports to every secondary school in the country – providing them fair and objective information on how they are performing.

Our tool could also be employed by the Education Review Office (ERO) or by the potential regional education hubs as recommended by the Tomorrow's Schools Taskforce report.

New Zealand needs fair and objective school evaluation. Without it, many top-performing low-decile schools will remain unrecognised, while many underperforming high-decile schools continue to fly under the radar.

The Initiative's school performance tool provides the missing piece of the New Zealand school evaluation puzzle – fair, objective and data-driven information.

Introduction

Secondary school education in New Zealand

The New Zealand secondary school education system is a black box. Students typically enter in year 9 (aged 12–13) bright-eyed and bushy-tailed; five years later, they leave in year 13 (aged 17–18) as mature young adults. But what happens inside schools is extremely complex and not easily understood.

This complexity is a product of a countless number of factors that contribute to a student's experience and accomplishments in school. Everything from a student's gender, ethnicity and native language to parental education, income and expectations have a hand in student achievement.¹ Great teachers and schools also have an undeniable role in student success.² Over time, all these factors influence the hundreds of assessments and exams that can lead to an National Certification of Educational Achievement (NCEA) qualification at the end of each senior year.³ Importantly, these qualifications, which are used to evaluate schools and create NCEA league tables, are the product of family background, teachers and school effects.

As a result of these entangled contributing factors, the Ministry of Education has struggled to identify and separate the contribution of schools from that of family background. Under current methods of school evaluation in New Zealand, schools that serve disadvantaged communities are more likely to receive poor reviews because of the different cohorts of students they serve.⁴ Consequently, some fantastic low-decile schools may not be recognised; at the same time, some underperforming high-decile schools might fly

under the radar. The Ministry can rank schools based on average NCEA performance within deciles; however, because of the significant number of students who attend schools out-of-zone, it is still not possible to get a clear picture of how schools are truly performing.⁵ Existing measures of school performance, including Education Review Office (ERO) reports and NCEA and University Entrance (UE) achievement rates – which only show absolute measures of student achievement – tell us as much about students and their parents as the quality of the school.

Objective, data-driven school evaluation is necessary to identify how schools are truly performing. Without it, schools are not evaluated on an even playing field.⁶ Determining which schools are high-performing and underperforming is imperative so the Ministry can learn from the better performing schools⁷ while providing support to those that need it.

Objective, data-driven school evaluation also empowers parents in deciding which school to send their children to. In many cases, this decision influences where they purchase or rent a home for their family. Homes in school zones where a school is perceived to be of higher quality earn a premium on the property market. For example, homes zoned for Epsom Girls Grammar School have the highest premium in the country, equal to 90.5%.⁸ Families without the means to move to the suburb of their choice may choose to drive hundreds of kilometres per year so their children can attend a school perceived to be of higher quality.⁹ Out-of-zone schooling is particularly prevalent in Christchurch, where students, in aggregate, travel 355,000 kilometres in one week, almost the entire

distance from the earth to the moon, to get to schools outside their school zone.¹⁰ For parents who are financially better off, a decision to send their children to a private school can cost more than \$100,000 over five or more years of schooling.¹¹

Without objective data, parents are forced to rely on anecdotal evidence and other unreliable proxies to inform their decisions.¹² For the past two decades, school quality has been inferred from decile ratings and league tables – *Stuff*'s school report and *The New Zealand Herald*'s Insights webpage.¹³ Together, these highly flawed proxies for school quality have fuelled decile drift and socioeconomic segregation in New Zealand secondary schools. In the 21 years since the decile funding model was introduced in 1995, the number of students in decile 8–10 schools has increased from 201,153 to 280,209; in contrast, the number of students in decile 1–3 schools has decreased from 188,089 to 179,929.¹⁴ During this period, 24% of all New Zealand students attended decile 1–3 schools compared with 45% of Māori students and 60% of Pasifika students.¹⁵ Even when low-decile schools are marked in the high-performing category by ERO, some families bypass them and enrol their children in average-performing high-decile schools outside their school zone.¹⁶ New Zealand parents are hardly to blame. Without other reliable sources of information on school quality, parents are left with no other choice but to resort to poor proxies and anecdotal evidence.

For far too long, New Zealanders have been left in the dark about fair and objective school evaluation. This has created a breeding ground for misconceptions, mistrust and isolation. What we currently know about our schools is inadequate, and New Zealand children are paying the price – particularly, those from the most disadvantaged communities. It is time principals, teachers and parents demanded better information about their schools from the Ministry of Education. We need to fix the way we evaluate our schools.

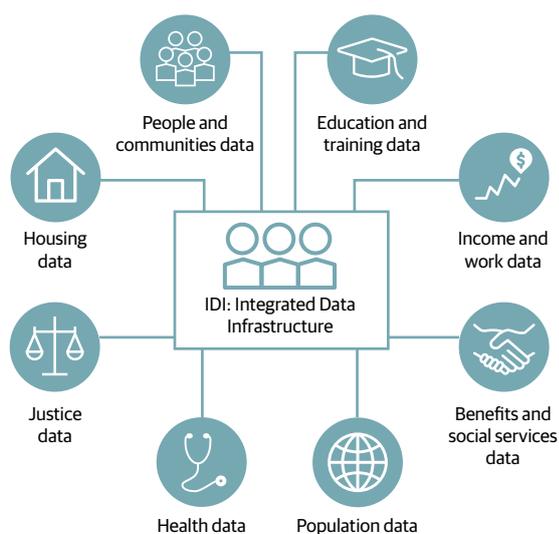
Fortunately, recent innovations in integrated data in New Zealand have given The New Zealand Initiative the opportunity to develop a school performance tool that can open the black box of education and examine what happens inside. Our tool provides a better picture of how much value each secondary school in New Zealand contributes to its students. Using the vast amounts of microdata in Statistics New Zealand's (SNZ) Integrated Data Infrastructure (IDI) – explained in greater detail in the following section – the Initiative has been able to separate the contribution of family socioeconomic background from the contribution of each school. This has enabled us to explain the differences in school performance typically seen in NCEA league tables. It has also enabled us for the first time in New Zealand to more fairly compare low- and high-decile schools, something we have not been able to do previously across all 480 secondary schools.¹⁷

The insights gained from the Initiative's school performance tool will be of value not only to principals, teachers and parents but also to the Ministry of Education, ERO and researchers. Information from our tool could be used to inform policy from the Ministry of Education, support school evaluation by ERO, and support management decisions by education hubs as recommended by the Tomorrow's Schools Taskforce report.¹⁸ If adopted by the Ministry, our tool could open the black box of education and provide the missing piece of the school evaluation puzzle – fair, objective and data-driven information.

Integrated Data Infrastructure

In 2011, SNZ started integrating data from all the government agencies in New Zealand in what is now called the IDI. Today, the IDI is New Zealand's largest research database. It contains data on more than 5 million people covering education, health, tax and income, social services, housing, and much more.¹⁹

Figure 1: An overview of the data available in the IDI



Source: Statistics New Zealand, “Integrated Data Infrastructure,” Website.

Researchers in government agencies, universities and think tanks such as the Initiative use this data to study New Zealand’s current and past population and investigate the impact of social services and government policies. Insights gained from the IDI can then be used to inform policy, including health interventions and education reform. Projects in the IDI are strictly limited to studies that will benefit the public and have no commercial gain.²⁰

Integrated data – and the IDI – is unique because data on individuals is de-identified while also linked across different ministries, social surveys, and censuses. This means Joe’s income can be linked to his education, his health, and his household during the last census. It also means Joe’s data is completely confidential; researchers can only see a unique ID number, never Joe’s name or address. Because of the sensitive and personal data in the IDI, SNZ highly prioritises confidentiality. To safeguard against any misuse of data, only qualified and vetted individuals and organisations are granted access to the data. Access is granted on a project-by-project basis, and only to the specific databases the project requires.

Over the past year, the Initiative has combined data from the Ministry of Education, New Zealand Police, Ministry of Social Development, Department of Corrections, Ministry for Children (previously Child, Youth and Family), Immigration New Zealand, Inland Revenue, and the 2013 Census to build a custom student dataset in the IDI. This dataset contains data on every²¹ student who attempted a unit or achievement standard²² as part of an NCEA level 1, 2 or 3 qualification. Because of the extensive range and level of detail contained in the IDI, the dataset had enough information to create a comprehensive snapshot of each student’s family background during the time he or she was at school. It includes information on each student’s ethnicity; the number and type of abuse events recorded by the Ministry for Children; the number of times that student was stood down, suspended or expelled; the number of times that student transferred secondary school; as well as whether the student was a refugee, had a disability, or had access to heat and the internet at home.

In addition to detailed information on each student, a comprehensive picture of parental socioeconomic background was linked to each student. This included information on home ownership, relationship/divorced status, education, income, benefit, police offence, and prison history.

In the final dataset, information on nearly 400,000 students and 480 schools over 10 years (2008–17) was used to investigate school effectiveness in New Zealand. This report discusses the results as follows.

Chapter 1 discusses value-added models and how other countries evaluate schools using them. Chapter 2 explains the Initiative’s school performance tool and how it evaluates school effectiveness. Chapter 3 discusses the results from our tool. Chapter 4 discusses the policy applications of the tool. Chapter 5 concludes.

Box 1: How deciles are calculated

Since the implementation of the decile funding model in 1995, school decile ratings have commonly been used in New Zealand media and research to discuss the differences between schools of different socioeconomic communities. While many New Zealanders know that decile is used for school funding, there are sometimes misconceptions about how they are calculated. The methodology is explained below.

Deciles are based on five equally weighted socioeconomic indicators:

1. **Household income:** The percentage of households with equivalent income in the lowest 20%, nationally adjusted for the number of adults and children in the household and age of the children.
2. **Occupation:** The percentage of employed parents in occupations who are at skill levels 4 or 5, according to the Australian and New Zealand Standard Classification of Occupations.
3. **Household crowding:** The percentage of households with an equivalised crowding index greater than 1.
4. **Educational qualification:** The percentage of parents with no tertiary or school qualifications.
5. **Income support:** The percentage of parents who directly (not as a partner) received

Jobseeker Support, Sole Parent Support, or Supported Living Payments (previously known as the Unemployment Benefit, Domestic Purpose Benefit, and Sickness and Invalid's Benefit, respectively) in previous years.

Source: Ministry of Education, "School deciles: How deciles are calculated," Education.govt.nz.

An important distinction between the decile system and our research presented in this report is the decile system uses area-level data while our school performance tool uses student-level data. In other words, deciles are calculated based on the households in the area the school is located, not the students who attend the school. This is not an issue for schools that have students predominantly coming from within that community; however, some schools have a large number of students from out-of-zone or outside that community.

In comparison, the calculations in our school performance tool are based on the students who attend the school, including students from both within and out-of-zone. Using student-level data leads to more precise school estimates and insights compared with only area-level data.

CHAPTER 1

Value-added models

The solution to school evaluation - Lessons from overseas

The challenge of developing a fair and objective school evaluation tool is not new or restricted to New Zealand; other countries, including the United States, the United Kingdom and Australia, also face this ongoing issue.²³

Principals, teachers, researchers and education professionals around the world have known for decades that family background matters greatly in student success.²⁴ Over the past four decades, a wide body of research has repeatedly demonstrated that the large differences in school league tables are predominantly a result of differences in school cohorts rather than differences in school effectiveness or quality.²⁵ In his meta-analysis, education researcher Robert Marzano found that schools only account for on average 20% of the variation in student achievement.²⁶

Subsequently, over time, countries and their relevant ministries/departments of education have developed various methods and tools to assess and evaluate schools objectively. The most common tool is the value-added model.

Unlike standard methods of assessment, which only look at snapshots of student achievement throughout the year, value-added models measure progress from the beginning to the end of the year.²⁷ The benefit of looking at progress rather than absolute achievement is value-added models acknowledge and adjust for the different levels of human capital that students bring with them to school.²⁸ Adjusting for different levels of human capital is a crucial component of value-added models as students come to school with a wide range of

abilities and backgrounds. Some students have excellent mathematics skills; others have a high aptitude for reading and writing; and many students, particularly those from disadvantaged communities, arrive with lower proficiencies in reading, writing and mathematics. Without acknowledging these differences, lower decile schools, on average, start the evaluation process one step behind – just as the students do when they start school.

The main purpose then of a value-added model is to provide the relevant ministry/department of education a tool that can impartially compare different schools that serve different cohorts of students. In other words, value-added models separate the contribution of the school from the contribution of each student's family background.

While the term value-added is commonly used across many countries, there is no universally accepted definition. The OECD defines it as “the contribution of a school to a student's progress towards stated or prescribed education objectives over time.”²⁹ Additionally, the OECD defines value-added modelling as “a category of statistical models that uses student achievement data to measure students learning gain.”³⁰

The many countries that have implemented value-added models typically seek to answer one if not all of the following three questions:³¹

1. What proportion of student achievement can be attributed to the school, the teacher or the family?
2. How effective is an individual school compared with other schools?
3. Which characteristics or institutional practices are associated with effective schools?

In each country, the exact statistical model varies in purpose, stakes and complexity. The exact model is dependent on the political climate and the available data in each country.

In some countries, value-added models were developed to evaluate a specific government education policy;³² in others, they were developed purely as a tool to evaluate school and teacher performance.³³

The stakes of each value-added model vary from relatively low stakes such as informing school and teacher improvement³⁴ to relatively high stakes such as teacher remuneration, employment and in some cases dismissal.³⁵

Data availability is, of course, crucial to the correct and unbiased implementation of a value-added model.³⁶ In some countries, only the results from standardised tests are used to inform value-added scores; in others, additional administrative data on student background is also used (contextualised value-added models). In the most complex models, students are linked to both their teachers and their school for every year they were enrolled. The New Zealand Initiative's school performance tool, described in greater detail in Chapter 2, builds and improves upon these existing models.

Importantly, whenever value-added models are implemented, they are typically not the only tool used to evaluate schools. Like New Zealand's ERO reviews, government-run qualitative evaluations that look at school culture, music and sports are also used in conjunction with the quantitative (data-driven) value-added evaluation.³⁷ The weights allocated to each evaluation tool vary from country to country, too.

Overview of value-added models

Value-added models can be broadly grouped into the following three categories:

1. gain score models
2. covariate adjustment models, and
3. multilevel models.³⁸

Gain score models

Gain score models are the simplest of the three models – only requiring the past and present results of a student to calculate the value-added scores for every school. A key assumption of this type of model is that the effect of family background on a student's academic results has been fully accounted for in their prior test scores, and that these factors do not have any future impact on student growth. This assumption is not always met, however, where it has been shown that prior performance *does* influence the rate of later learning.³⁹ As a result, gain-score models can produce biased school estimates that are less accurate than the ones produced from the more complex, covariate adjustment model and multilevel model.⁴⁰ Nonetheless, the key strength of this type of model is its simplicity and ease of implementation, hence its widespread use in many countries.

Gain score models are being used in the 'My School' student gain reporting in Australia⁴¹ and the Tennessee Value-Added Assessment System (TVAAS) in the United States.⁴² In each case, the purpose of the model varies slightly in addition to the exact gain score model used. In Australia, the My School website, www.myschool.edu.au, aims to give parents quality information about how their local schools are performing.⁴³ In Tennessee, the main purpose of the TVAAS is to evaluate the impact of the *Education Improvement Act*, which increased education funding significantly in the state.⁴⁴

Covariate adjustment models

Covariate adjustment models improve upon gain score models by including additional information on student background characteristics. By including this additional information, covariate adjustment models are typically more robust (reliable) and accurate.

In contrast to gain score models, which only use the observed scores of each student to calculate school value-added estimates or scores, covariate adjustment models use simple statistical modelling to predict average school performance, given the characteristics of each school's students. In these types of models, the value-added scores for each school are calculated using the difference between the observed and predicted mean school performance.

The main drawback is covariate adjustment models do not take into account classroom (teacher) level effects; in other words, they only calculate the value-added scores for schools, not teachers. As a result, the estimated school value-added scores can be less robust compared to more complex multilevel models.⁴⁵ One example of a covariate adjustment model is the Dallas Value-Added Assessment System (DVAAS) in the United States.⁴⁶

Multilevel models

Multilevel models are the most complex and sophisticated among all three categories of value-added models. Using administrative data on students, teachers and the school, multilevel models improve upon the covariate adjustment model by including teacher-level effects in addition to school-level effects. This is done using a two-step or a three-step statistical model. In a three-step model, the first step is at the student level, the second at the teacher level, and the third at the school level.⁴⁷ The result is school value-added estimates that are more robust and accurate. Unlike the previous two models, multilevel models can distinguish between high value-add schools and high value-add teachers. This can be useful for determining what proportion of high-performance or poor-performance can be attributed to school-specific or teacher-specific reasons.⁴⁸ The downside to multilevel models is their complexity and the amount of data required. Multilevel value-added models have been implemented in New South Wales, Australia;⁴⁹ the United Kingdom;⁵⁰

Hong Kong;⁵¹ and several states/districts in the United States.⁵²

Robustness (reliability) across models

The level of robustness in school (or teacher) value-added scores depends on the value-added model used and the quality of data collected. The preferred level of robustness and model depends on data availability and the purpose of the research. Those decisions may be made by educators and policymakers. The higher the stakes, the greater is the need for robust findings. Every model has different strengths and weaknesses, and different costs and benefits. None are perfect.⁵³ When assessing accountability for decisions, allowance must be made for model weaknesses and imprecisions in estimated value-added scores. Decision-making must be judicious.

Which value-added model should New Zealand use?

Multilevel value-added models are an obvious choice, being the most robust and accurate. However, as described earlier, the political climate, availability of data, and purpose always play a part in choosing a model. Multilevel models are not always the preferred choice, even when they can be implemented.

In New Zealand, we do not have centrally located data that can link students to their teachers, nor do we have a political climate that demands a systematic evaluation of teachers. Additionally, New Zealand does not have standardised testing to develop a standard value-added model. External evaluation tools such as Value-Added VA-12 and VA-13 reporting⁵⁴ are available to New Zealand schools privately; however, these services are mostly accessed only by high-decile schools because of the additional cost.⁵⁵

These hurdles should not prevent researchers from developing better and fairer measures of school performance.⁶ In fact, New Zealand excels in other ways – SNZ’s IDI is a world-leading integrated database, even besting countries like the United States and Australia in integrated data. By leveraging the data in the IDI, New Zealand could overcome hurdles and build better and fairer measures of school performance than we currently have.

The New Zealand Initiative has done just that. Following a year-long analysis of New Zealand’s education and socioeconomic data, we have developed New Zealand’s first contextualised value-added model. While our model differs from existing value-added models in many ways, it answers the same questions. Chapter 2 goes into more detail.

CHAPTER 2

School performance tool

The Initiative's school performance tool

The overarching goal of the Initiative's school performance tool is to help decision-makers lift the performance of New Zealand's education system; however, the way it contributes is through several smaller objectives. Like the value-added models described in Chapter 1, the Initiative's school performance tool attempts to answer the following three questions:

1. What proportion of student achievement can be attributed to the school and to factors outside the control of schools?
2. How effective is each secondary school compared with the other 479 secondary schools in New Zealand?⁵⁷
3. Which characteristics or institutional practices are associated with effective and ineffective schools?

While the Initiative's school performance tool shares similar goals to value-added models used overseas, the way it approaches these objectives is slightly different. One of the key requirements needed for a value-added model is standardised testing; without it, you cannot create a value-added model as described in the literature.

Unfortunately, New Zealand does not have standardised testing. Currently, New Zealand only has NCEA as its main secondary school qualification.⁵⁸ Students can complete an NCEA level 1, 2 or 3 qualification via an astronomical number of combinations of unit and achievement standards across several different subjects and years.⁵⁹ While NCEA gives students extreme flexibility, it means one student's 80 level 1 credits are rarely comparable with another student's

80 level 1 credits. Additionally, it means one student's 80 level 1 credits are not comparable with that same student's 80 level 2 credits. Consequently, we cannot create a 'true' value-added measure in New Zealand.⁶⁰

Fortunately, there is another way. Using the vast amounts of information in the IDI, the Initiative has been able to create a variation of the value-added model, alternatively termed a contextualised attainment model, or simply, a contextualised value-added model. Instead of using prior test scores to adjust for differences in family background, the Initiative has used the vast amounts of microdata in the IDI.

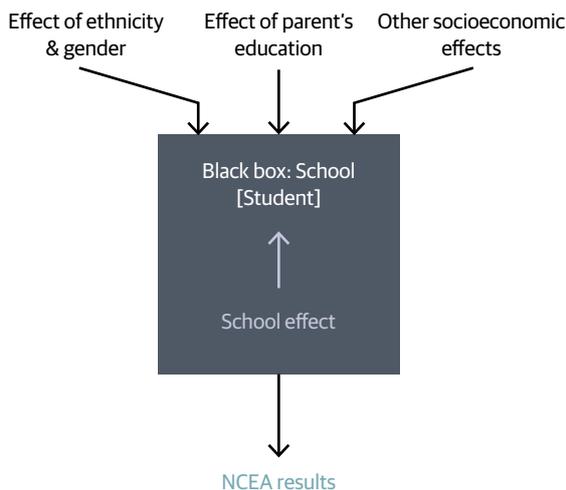
Furthermore, considering the issues with NCEA outlined above, the Initiative developed two NCEA derived metrics – i) a Weighted Relative Performance Index (WRPI) score, and ii) a weighted NCEA score – supplemented by a third NCEA derived metric, an expected percentile score developed by Michael Johnston at the New Zealand Qualifications Authority (NZQA). Both Initiative-NCEA derived metrics are discussed further in the Appendix and in a previous Initiative report, *Score! Transforming NCEA data*.

Using this NCEA data in combination with data from several Ministry and Census databases in the IDI, our school performance tool has separated the contribution of a student's family background from that of the school. While the method for separating the effect of family socioeconomic background differs from conventional value-added models, the same main objectives are reached.⁶¹

How the Initiative's school performance tool works

The Initiative's school performance tool achieves its main objectives in somewhat the opposite way to the education black box. In an education black box, various factors – such as the influence of parental education, ethnicity, gender, and home environment, in addition to other socioeconomic factors – contribute to a student's grades in school. What comes out at the end of school – in New Zealand's case, NCEA results – is a combination of both the school effect and the student's family socioeconomic background effect. The education black box is illustrated below (see Figure 2).⁶²

Figure 2: Education black box

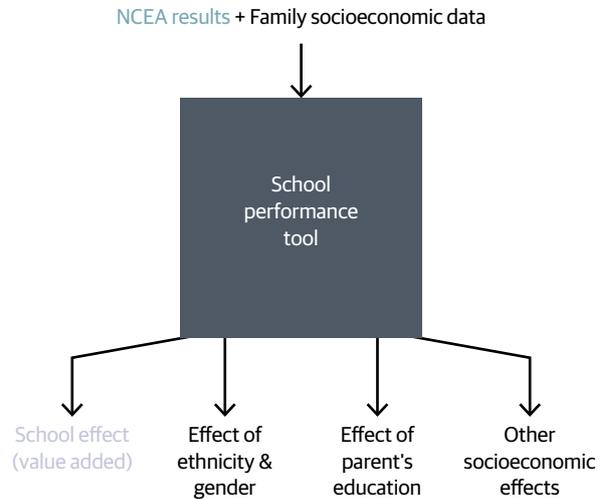


The Initiative's school performance tool works in the opposite way. Instead of NCEA results exiting the black box as the product of family background and school contribution, as shown in Figure 2, NCEA results enter our school performance tool in the beginning, while estimates for the individual contributing factors exit the tool at the end.

Using the model, we have been able to estimate individual effects of ethnicity, gender and parental education, in addition to various other socioeconomic effects. Importantly, the model has

also estimated individual school effects – in other words, (contextualised) value-added scores for every secondary school in the country (see Figure 3).⁶³

Figure 3: School performance tool



While the industry term for the Initiative's school performance tool is a contextualised value-added model, the technical term is a fixed-effects model with least squares dummy variables (LSDV) estimators.⁶⁴ The fixed effects are the individual secondary school effects estimated using the LSDVs. However, for the remainder of this report, we refer to our model as our school performance tool.

The full technical details of the model's development are discussed in the corresponding technical report, *Separating School and Family: Evaluating the effects of school and family background on student performance in NCEA*.

Data from the Integrated Data Infrastructure

In constructing the Initiative's school performance tool, we combined data from the Ministry of Education, New Zealand Police, Ministry of Social Development, Department of Corrections, Ministry for Children (previously Child, Youth and Family), Immigration New Zealand,

Inland Revenue, and the 2013 Census in the IDI. After merging all the Ministry and Census databases, the final dataset contained information on nearly 400,000 students in all 480 secondary schools.⁶⁵ The dataset covered the 10 years from 2008 to 2017. Summary statistics of our dataset are available in *Separating School and Family*.

Technical details of the Initiative's school performance tool

While our school performance tool looks complicated, it can easily be broken down into eight main components illustrated in Equations 1 and 2. Note that for each term in both equations – Y_i , T_i , X_i , Z_i , D_i , ϵ_i – i refers to one student; this is because we have used student-level data.

Equation 1: School performance tool annotated

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 W_i + \beta_4 Z_i + \beta_5 D_i + \epsilon_i$$

NCEA outcome
Time effects
Parent effects
School value-added score/ effect
Random error

Student constant
Student effects
School type effects

Equation 2: School performance tool unannotated

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 W_i + \beta_4 Z_i + \beta_5 D_i + \epsilon_i$$

Model components

1. **NCEA outcome:** The first component of our model is the NCEA outcome variable of interest term, denoted as Y_i . When evaluating schools, we must decide what outcome or outcomes to evaluate schools on. The obvious first choice and the choice most often used in school evaluation is student academic outcomes.

While we acknowledge that schools do more than just teach their students English, mathematics, science, and various other subjects, roughly indicated by NCEA achievement, it is difficult and in many cases impossible to evaluate schools on more qualitative measures such as student wellbeing

and performance in music and sports using a model like ours.⁶⁶ Fortunately, ERO reviews are capable of measuring qualitative outcomes.

In the first run of our school performance tool, we evaluated schools on 10 academic outcomes – all derived from each student's NCEA results. These variables ranged from the Initiative's WRPI score, weighted NCEA score, expected percentile score, and University Entrance (UE) achievement. Table 1 lists all the 10 variables we used.

Importantly, in future reports we will evaluate schools on later life outcomes such as NEET (not in employment, education, or training) status one, three and five years after college; progression; and completion of tertiary education after college – in addition to several other later life outcomes. This is particularly important as some schools have a large proportion of students attempting the Cambridge International General Certificate of Education or the International Baccalaureate qualification.

Table 1: Student NCEA outcomes variables

Student NCEA outcome variables	
1.	WRPI NCEA level 1 score
2.	WRPI NCEA level 2 score
3.	WRPI NCEA level 3 score
4.	Expected percentile NCEA level 1 score
5.	Expected percentile NCEA level 2 score
6.	Expected percentile NCEA level 3 score
7.	Weighted NCEA level 1 score
8.	Weighted NCEA level 2 score
9.	Weighted NCEA level 3 score
10.	University Entrance

2. **Student constant:** The second component of our model is the student constant term, denoted as β_0 . The student constant can be thought of as the number of NCEA credits/NCEA score we expect a student to earn independent of family socioeconomic background and the school effect.

Box 2: Interpreting estimated effects

When interpreting the estimated effects from our model, particularly the results in Tables 2 and 3, it is imperative they are *not* interpreted as causal effects. Our model does not propose that any of our estimated effects *cause* a student to improve or worsen their NCEA grade. The estimated 'effects' are more appropriately termed 'associations' or 'correlations'; however, it is common in the literature to refer to regression estimates as 'effects' – as we have done in this report.

Example 1: β_{female} is the estimated average performance of female students compared with male students in NCEA level 1, 2 or 3, after separating the effects of the other background characteristics listed in Tables 2 to 4. Positive β_{female} results indicate that on average, female students perform better than male students in NCEA level 1, 2 or 3.

Example 2: $\beta_{māori}$ is the estimated average performance of Māori students compared with non-Māori students after separating the effects of the other background characteristics listed in Tables 2 to 4. If $\beta_{māori}$ has a negative estimated effect on a student's NCEA results, it does not mean a Māori student *will* perform worse compared with a non-Māori student. It does, however, indicate that Māori students perform worse on average compared with

non-Māori students in NCEA even after adjusting for the (other) background characteristics included in our model.

Example 3: $\beta_{mother's\ education}$ is the estimated effect of a mother's education on a student's NCEA level 1, 2 or 3 achievement – in particular, the effect of mother's education by level of education attainment (none, high-school certificate, diploma, bachelor's degree, or post-graduate degree). If $\beta_{mother's\ education}$ is increasing by level of education attainment, it does not mean a mother with a PhD ensures her child will do well in NCEA; it does mean, however, that students with mothers with PhDs perform better in NCEA, on average, even after adjusting for other family background characteristics included in our model.

When interpreting these results, the estimated effects from student and family background characteristics are not about profiling *individual* students; it is about acknowledging them and then adjusting for these background characteristics we know affect a student's academic outcomes *on average* to get a better measure of individual school performance. These results are not the focus of the model or this report; however, important lessons can still be learned.

- Time effect:** The third component of the model is the time effect term, denoted as $\beta_1 T_i$. Specifically, T_i is an indicator variable for the year the student sat NCEA level 1, 2 and 3, and β_1 is the estimated effect of that NCEA year. This component is important because our dataset contains pooled NCEA data over the 10 years from 2008 to 2017. Different NCEA years have different cohorts of students, each of which may perform differently from the previous and the next year. Additionally, there may be nuances within NCEA that change from year to

year, possibly affecting NCEA results. This component accounts for these differences. There may be several reasons for these differences; however, the explanation for them is beyond the scope of this report.

- Student effects:** The fourth component of our model is the student background effect term, denoted as $\beta_2 X_i$. Specifically, β_2 is a range of individual estimated effects from a range of student background characteristics indicated by X_i . Table 2 lists all the 14 background characteristics included in

the model. Box 2 explains how to interpret the various estimated results.

Lastly, the variables included in Tables 2 to 4 are not necessarily the final variables that would or should be used in a practical, implementable version of this model. Our model is very much a first run where we went with a ‘kitchen sink’ approach on including family socioeconomic background variables. Analysis of our results will be useful for informing which variables should be retained and which should be removed in future iterations of our model. Further analysis and discussion of which variables should and should not be included is available in the corresponding technical report, *Separating School and Family*.

Table 2: Student background characteristics variables

X_i : Student background characteristics variables	
1.	Female (Y/N)
2.	Ethnicity Māori; Pasifika; Australian; Asian; European; Middle Eastern; Latin American; African
3.	Number of abuse events by category identified by CYF Sexual abuse; Physical abuse; Emotional abuse; Neglect abuse; Self-harm abuse; Behavioural abuse
4.	Refugee (Y/N)
5.	Disability (Y/N)
6.	English as a second or other language (ESOL) (Y/N)
7.	Reading recovery (Y/N)
8.	Number of suspensions ⁶⁷
9.	Number of stand downs
10.	Expulsion (Y/N)
11.	Number of secondary schools attended
12.	Percentage of internal credits by NCEA year NCEA level 1, NCEA level 2, NCEA level 3
13.	Access to the internet at home (Y/N)
14.	Access to heat at home (Y/N)

5. **Parental effects:** The fifth component of our model is the parental background effect term, denoted as $\beta_3 W_i$. Specifically, β_3 is a range of individual estimated effects from a range of

parental background characteristics indicated by W_i . Table 3 lists all 12 background characteristics included in the model.

As noted earlier, parental influence plays a role in how well a student performs academically. This component of the model acknowledges parental influence and adjusts for it.

The interpretation of the parental background characteristic variables is similar to the interpretation of the student background characteristic variables described earlier, where the individual family background characteristics estimates are only predictive of a student’s academic performance in NCEA.

Finally, note that two variables – ‘Mother’s log income’ and ‘Father’s log income’ – are the logarithm of each student’s mother’s and father’s average income from 2000 to 2017. We have used log income to help interpret results in the technical report; it does not alter the school value-added estimates.

Table 3: Parental background characteristics variables

W_i : Parents’ background characteristics variables	
1.	Parents’ home ownership (Y/N)
2.	Parents divorced (Y/N)
3.	Mother’s education None; High school certificate; Diploma (level 4–6); Bachelor’s degree (level 7); Post-graduate degree (Master’s/PhD)
4.	Father’s education None; High school certificate; Diploma (level 4–6); Bachelor’s degree (level 7); Post-graduate degree (Master’s/PhD)
5.	Mother’s log income
6.	Father’s log income
7.	Mother’s benefit spell (weeks)
8.	Father’s benefit spell (weeks)
9.	Number of mother’s offences
10.	Number of father’s offences
11.	Mother has interacted with New Zealand Corrections (Y/N)
12.	Father has interacted with New Zealand Corrections (Y/N)

6. **School type effects:** The sixth component of our model is the school type effect term, denoted as $\beta_4 Z_i$. Specifically, β_4 is the range of individual estimated effects from various school types indicated by Z_i . Table 4 lists all four school types included in the model.⁶⁸ The variables in Table 4 indicate the average estimated effect of various school types, including whether the school was girls-only, boys-only, state, private or state-integrated. Finally, it included the isolation index of each secondary school. These variables reveal categorical differences across the different school types in New Zealand.

Table 4: School type variables

Z_i : School type	
1.	Girls only school (Y/N)
2.	Boys only school (Y/N)
3.	State school (Y/N)
4.	School isolation index

7. **School value-added score/effect:** The seventh and most important component of our model is the individual school value-added score term, denoted as $\beta_5 D_i$. Specifically, β_5 is a range of individual school estimates, while D_i is a range of indicator variables for every secondary school in our

final dataset. In other words, this component of the model indicates the “value-added” contribution of every secondary school in New Zealand. As noted earlier, this component measures fixed effects, where the school indicator variables denoted as D_i are the LSDVs.

Importantly, while this is the focus of our model, SNZ’s confidentiality rules prevent us from publishing the results from each school. Chapter 3 goes into further detail on how we present the results in a way that complies with SNZ’s confidentiality rules and in an unbiased matter.

8. **Random error:** The eighth and final component of our model is the random error term, denoted as ϵ_i . This component captures the ‘raw ability’ of students – it is the part of their performance that remains after separating the effects of family background and the effects of the school.

In total, our school performance tool evaluated all 480 secondary schools on 10 different NCEA outcomes. Chapter 3 explains our results.

CHAPTER 3

School performance tool results

Summary of decile results

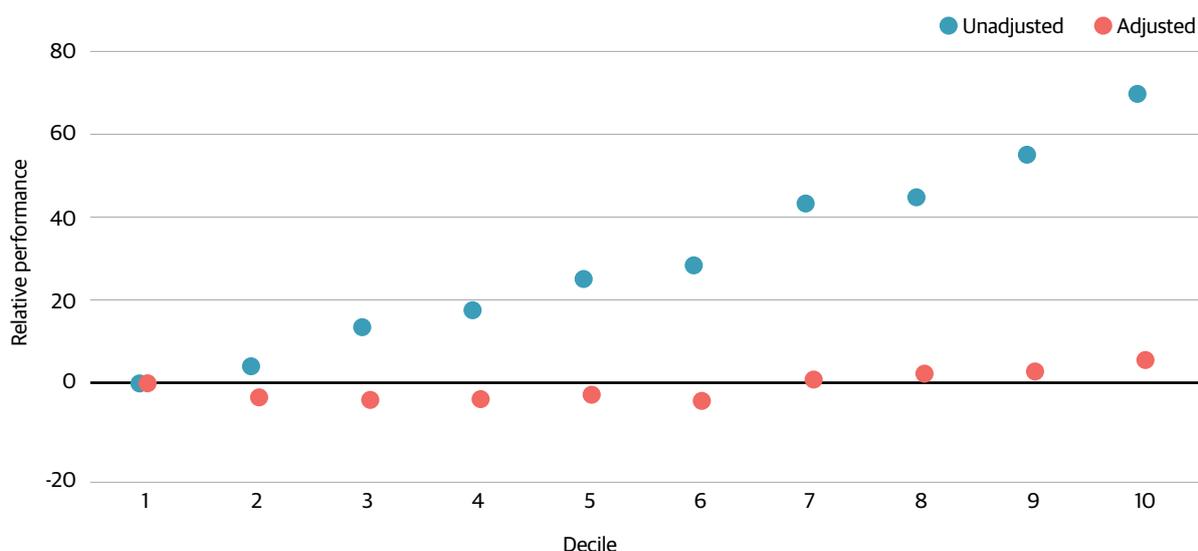
The results of our school performance tool have several implications for education in New Zealand. The first set of results, discussed in an earlier Initiative research note, *Tomorrow's Schools: Data and evidence*, presented a sample of our results on the average performance of schools across deciles.⁶⁹ In summary, *Tomorrow's Schools* provided the first empirical evidence against the pervasive myth that “decile is a proxy for school quality”. Figure 4 is one of the four figures discussed in *Tomorrow's Schools*.⁷⁰

In Figure 4, each decile (approximately 50 schools) is represented by two points, one unadjusted (blue) and one adjusted (red). The unadjusted scores show the average performance of schools within each decile, not adjusting for family socioeconomic background.⁷¹ In contrast,

the adjusted scores show the average performance of schools within each decile after adjusting for family socioeconomic background.

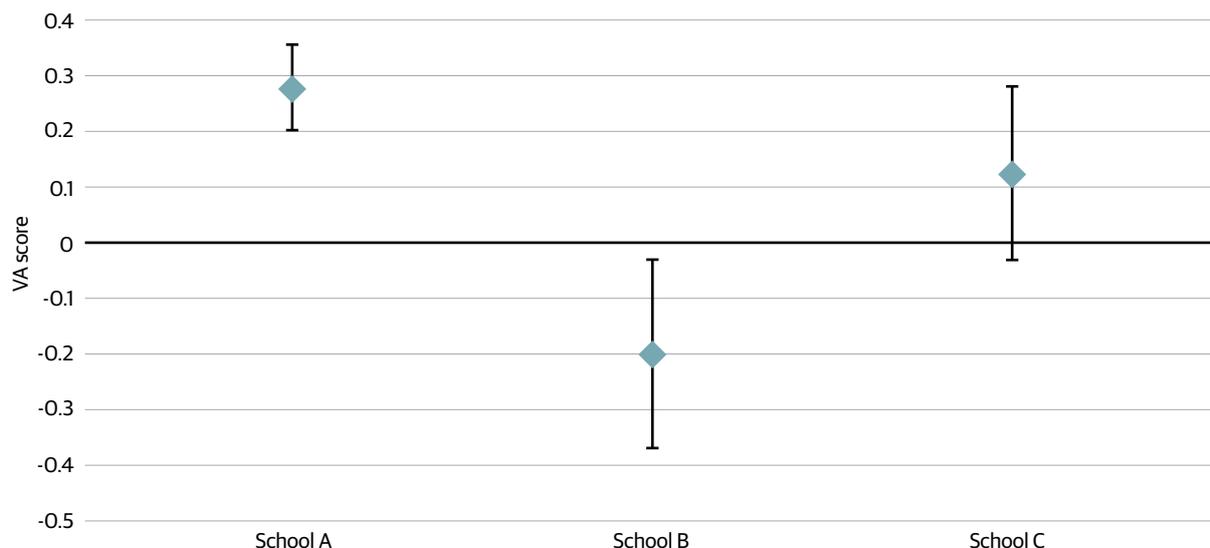
In both cases, each school was evaluated on the NCEA level 1 WRPI score of its students. When we broke down these results, we were not surprised to find that higher decile schools outperformed lower decile schools when our school performance tool did *not* adjust for each student's family background characteristics. This is reflected by the rising blue unadjusted points in Figure 4 that show relative performance increasing with school decile. This disparity in school performance replicates what we see in NCEA school league tables and also in the inequality in education outcomes that several international reports and assessments have shown in recent years.

Figure 4: Unadjusted and adjusted average performance of secondary schools within each decile based on each student's NCEA level 1 WRPI score (2008-17)



Source: Author's calculations from Statistics New Zealand's Integrated Data Infrastructure (IDI).

Figure 5: Hypothetical value-added scores for three schools



Source: Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools” (Sydney: Centre for Education Statistics and Evaluation, NSW Department of Education and Communities, 2014), Figure 4.

However, once our school performance tool adjusted for each student’s family background, the performance differences between schools of different deciles disappeared. This is reflected by the level red adjusted points in Figure 4 that show higher decile schools have approximately the same relative performance compared with decile 1 schools.⁷²

Put another way, our results show that the inequality in education outcomes between school deciles evident in school league tables is not a result of large differences in school quality, but rather large differences in family socioeconomic background, particularly differences in parental education.

However, as we alluded to in our *Tomorrow’s Schools* research note, Figure 4 is not the full story. The results in *Tomorrow’s Schools* only show the average performance of schools within and across deciles. Our school performance tool is much more powerful and accurate. By leveraging the datasets in the IDI, our school performance tool was able to evaluate every secondary school in New Zealand individually.

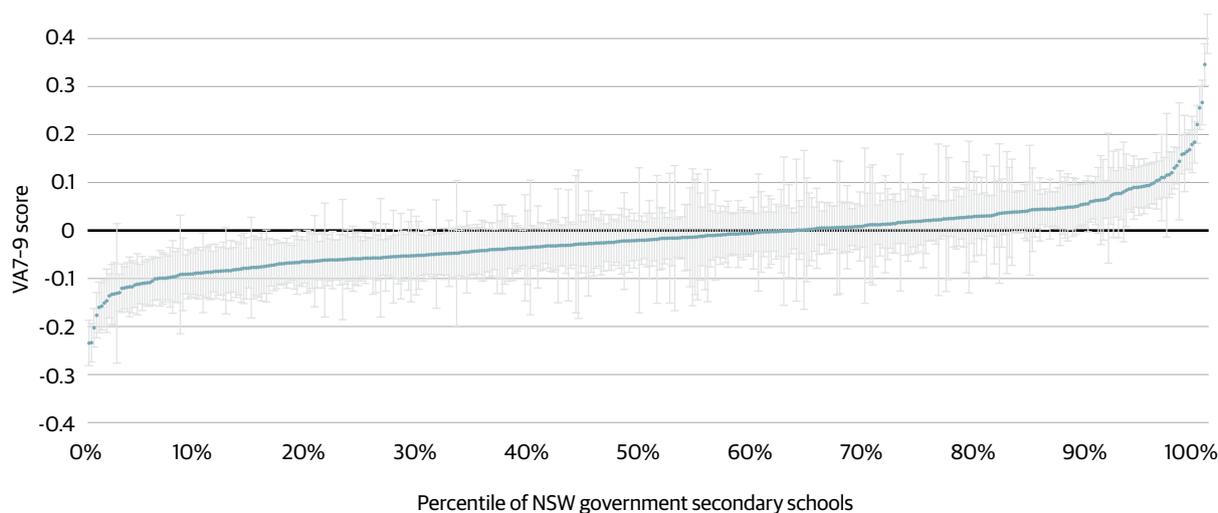
Individual school performance results

The results presented in this section tell more of the story. Figure 7 shows one of the many results from our school performance tool that illustrate the distribution of individual school performance.

Unfortunately, because of confidentiality rules set and enforced by SNZ, the results presented here are less intuitive than we would have liked. We would have liked to present the results in the form of Figure 5, where each secondary school was represented by one point representing the “value-added” score as calculated by our school performance tool. Additionally, each point would have a 95% confidence interval band that essentially represented how accurate our results are – where the larger the band the less accurate the results (see schools B and C) and the smaller the band the more accurate the results (see school A).⁷³ Around the world, this is typically how value-added results are presented.⁷⁴

Extrapolated to all the secondary schools in the country, this figure would look something like

Figure 6: NSW: Value-added scores for the entire school population



Source: Centre for Education Statistics and Evaluation, “Using Value-Added Measures to Identify School Contributions to Student Learning” (Sydney: 2014), Figure 4.

Figure 6, where again, each school is represented as one point with a 95% confidence interval band. To best demonstrate the distribution of school performance, the school “value-added” scores are ordered from smallest to largest. Figure 6 comes from a NSW school value-added model report.⁷⁵

Unfortunately, the confidentiality rules set and enforced by SNZ prevent the Initiative from producing figures where each school is represented by one point, even where each school is *not* identified by its name or unique ID number.⁷⁶ Our corresponding technical paper, *Separating School and Family*, discusses this issue in greater detail. We have shared our frustrations with the relevant government Ministers, who were receptive.

To overcome our confidentiality problem, we turned to a more creative graphic format. Like in Figure 6, we ordered all our school value-added scores from lowest to highest. However, in Figure 7 we also applied a locally weighted scatterplot smoothing (LOWESS)⁷⁷ curve to our school value-added scores to show the distribution of school performance while complying with the confidentiality rules enforced by SNZ.

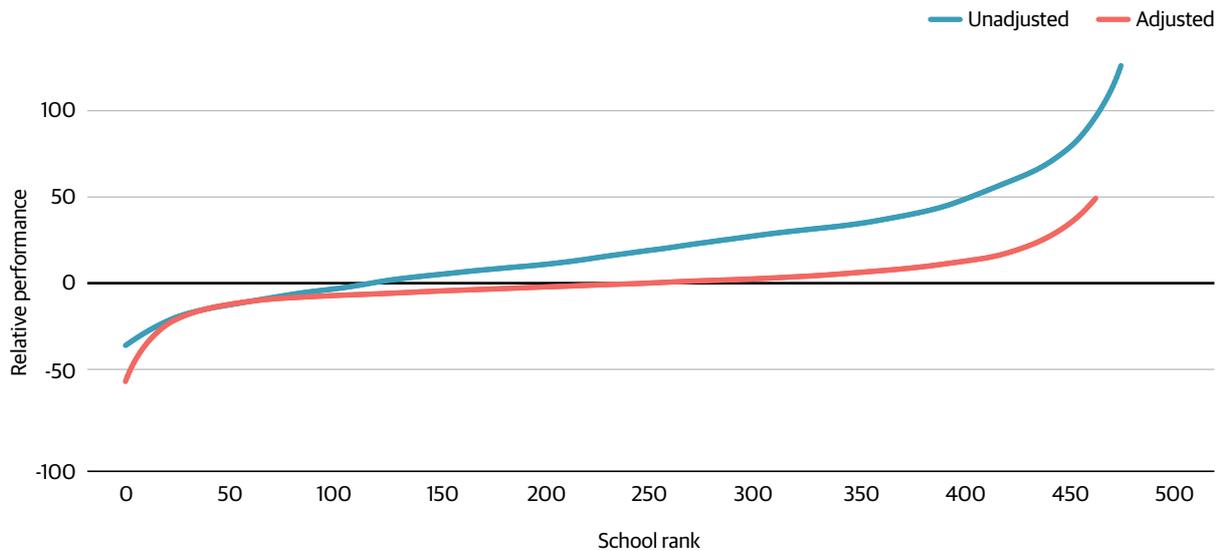
Unfortunately, as part of this process we were not able to add the 95% confidence interval bands. Fortunately, because the results are distributed tightly around the red baseline,⁷⁸ this has not distorted our results significantly.

In contrast to Figure 6, Figure 7 has two value-added scores for each school; the first is the unadjusted value-added scores represented by the blue LOWESS curve, while the second is the adjusted value-added scores represented by the red LOWESS curve. Like the decile results, the blue unadjusted scores are the value-added scores before adjusting for family background, while the red adjusted scores are the value-added scores after adjusting for family background.

As in Figure 4, the variation in school performance is much larger when differences in family background are not adjusted for; this reflects what we already see in school league tables.

However, once our school performance tool separated the contribution of the family background from the contribution of the school, the variation in school performance was much smaller: approximately 80% of schools perform almost identically when evaluated on their

Figure 7: Unadjusted and adjusted performance of New Zealand secondary schools based on each student's NCEA level 1 WRPI score (2008-17)



Source: Author's calculations from Statistics New Zealand's Integrated Data Infrastructure (IDI).

students' NCEA performance. While Figure 7 only shows the results from our NCEA level 1 WRPI score evaluation, the full set of results in Figures 11 to 20 in *Separating School and Family* show similar results for all 10 NCEA outcomes we evaluated schools on.

Importantly, while most schools perform very similarly, there are still outliers at the top and bottom of the distribution once our school performance tool separated the contribution of family background. In other words, there are still high-performing and low-performing secondary schools in New Zealand, where the top-performing school(s) score nearly one standard deviation higher than the median-performing school, and the bottom-performing school(s) score nearly one standard deviation below the median-performing school.

However, like Figure 4, Figure 7 does not tell us the full story. Figure 4 illustrates average school performance across deciles, while Figure 7 illustrates individual school performance across all schools. Figures 8 and 9 tell us more of the story.

Individual school performance across deciles

Figures 8 and 9 illustrate the distribution of individual school performance across deciles, where Figure 8 presents the unadjusted results and Figure 9 presents the adjusted results. Both figures show the results from the NCEA level 1 WRPI score evaluation.

Figures 8 and 9 were constructed in several steps. The first was ordering each school's value-added scores from lowest to highest, the same step that was done in Figure 7. Following this, we categorised the top 25% of schools as high-performing, the middle 50% as average-performing, and the bottom 25% as low-performing. Subsequently, schools in each category were tallied in tables by decile, and then shown as stacked bar graphs in Figures 8 and 9.⁷⁹ A summary of each performance category is shown in Table 5.

Table 5: Relevant colours and categories for Figures 8 and 9

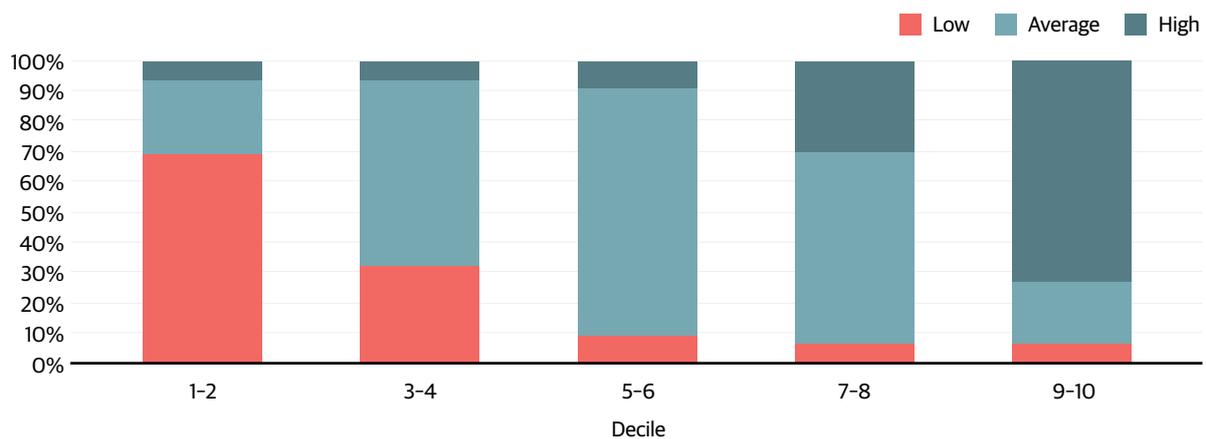
	Low	Bottom 25% of schools
	Average	Middle 50% of schools
	High	Top 25% of schools

We chose the performance band of 25% as a result of the strict microdata output rules set and enforced by SNZ.⁸⁰ Our preferred performance band was top 10% as high-performing, middle 80% as average-performing, and bottom 10% as low-performing, given the results presented in Figure 7; however, a performance band of 10% would have prevented us from illustrating the results as reliable and representative of the

‘true’ results calculated in the IDI data lab. Part 3 of the results section in *Separating School and Family* goes into this in greater detail.

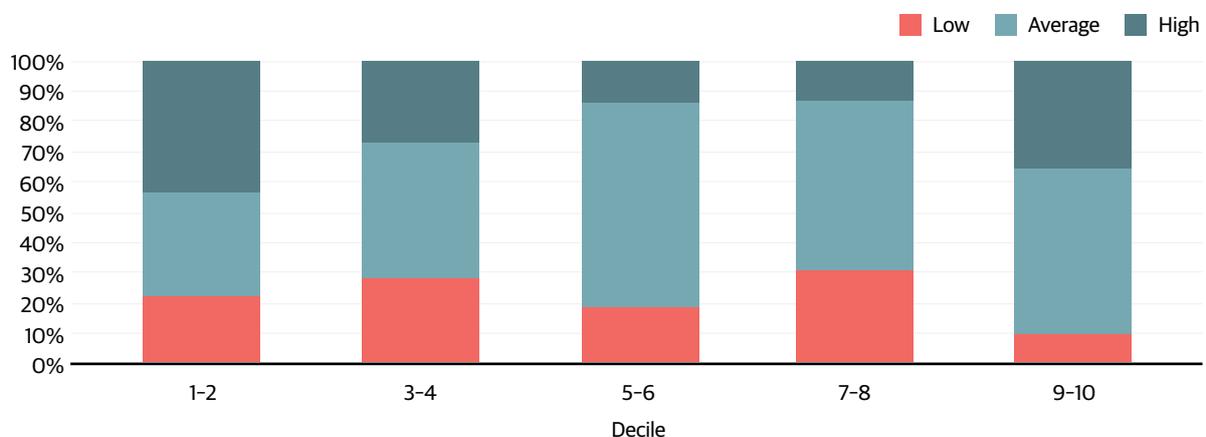
Figure 8 shows that high-decile schools dominate the high-performing category, while low-decile schools dominate the low-performing category when differences in family background had *not* been adjusted for by our school performance tool. Figure 8 illustrates 73% of decile 9 and 10 schools as high-performing compared with 69% of low-performing decile 1 and 2 schools when evaluated on University Entrance. Given the polarising results in Figure 8, it is not surprising that some parents have adopted the decile rating of a school as a proxy for school quality.

Figure 8: Distribution of school performance: University Entrance (unadjusted)



Source: Author’s calculations from Statistics New Zealand’s Integrated Data Infrastructure (IDI).

Figure 9: Distribution of school performance: University Entrance (adjusted)



Source: Author’s calculations from Statistics New Zealand’s Integrated Data Infrastructure (IDI).

However, once our school performance tool adjusted for differences in family background, the results were more balanced. Figure 9 shows particularly large proportions of high-performing schools in both high- and low-decile schools; specifically, 44% of decile 1 and 2 schools as high-performing compared with 36% of decile 9 and 10 schools as low-performing. In particular, when evaluated on University Entrance, 42 decile 1 and 2 schools outperform 75% of every other secondary school in the country; in contrast, 9 decile 9 and 10 schools are in the bottom 25% of all secondary schools in the country.

In particular, when evaluated on University Entrance, 42 decile 1 and 2 schools outperform 75% of every other secondary school in the country

Figures 8 and 9 are only two of the 20 figures that demonstrate the distribution of school performance across deciles. All 20 figures are illustrated and discussed further in our technical report, *Separating School and Family*. Importantly, out of the 10 NCEA outcomes we evaluated schools on, seven show almost identical

results to those shown in Figures 8 and 9. In three out of the 10 results, high-performing schools still dominate in higher decile schools; however, high-performing schools are still present in both low- and middle-decile schools. Importantly, the number of high-performing schools is more evenly distributed after our school performance tool has separated the contribution of family background. This result is consistent across all our 10 NCEA outcome measures.

Future of our school performance tool

In this chapter, we have unveiled a more accurate distribution of school performance in New Zealand. We have shown that most schools perform very similarly once our school performance tool separated the contribution of family background. Additionally, we have shown that there are outliers among New Zealand secondary schools, and there are still high-performing and low-performing schools once our tool separated the contribution of family background.

Box 3: Hypothesis explaining the distribution of school performance

Interestingly, when evaluated on University Entrance (and expected percentile), middle-decile schools in our model fared the worst. Only 14% of decile 5 and 6 schools were in the high-performing category compared with 44% in deciles 1 and 2 and 36% in deciles 9 and 10 (see Figure 9).

One hypothesis from one of our referees on why low- and high-decile schools have a higher proportion of high performers is what are better teachers attracted to. Better teachers may be more attracted to high-decile schools for a more 'comfy' and challenging life with more able students; in contrast, better teachers may also be attracted to low-decile schools because they believe in their potential to make a difference to a child's life.

Crucially, this is just speculation. It also reveals one major limitation of our model, that is, it is unable to distinguish between school-specific and teacher-specific effects. This was alluded to in Chapter 1, where we described the major strength of multilevel models as the ability to distinguish between school-specific and teacher-specific-effects.

To determine the drivers of these differences in school performance, the Ministry of Education would need to study these high-performing schools via in-school visits – similar to ERO reviews. As we noted earlier, each value-added model is in some way imperfect – this is one limitation of our model.

Some of these schools should be celebrated and learned from while others may need additional support. We have demonstrated what can be done with the IDI. However, we are limited by what we can do with the data and with our tool.

Our tool is limited by who wields it and by the rules set and enforced by SNZ. The Ministry of Education could use our tool for the better and open the black box of education. The Ministry could also identify which schools are the star performers and which schools need additional support; whether the Ministry chooses to make that information public is up to them.

This chapter has shown what the Initiative's tool can do. Chapter 4 discusses how our tool can be implemented through policy.

CHAPTER 4

Policy applications

The previous chapter outlined what the Initiative's school performance tool could do. However, the tool's potential can only be reached if it is implemented correctly. Around the world, successful value-added models are implemented by the relevant ministry/department of education. The same should be done here.

Complement to ERO

As described in previous chapters, our school performance tool is a more objective and fairer way to measure school performance. However, many aspects of school performance cannot be evaluated by our school performance tool. For example, qualitative measures such as school culture, music and sporting activities cannot be evaluated by a technical model like ours. Therefore, the best way to implement our tool is in conjunction with complementary qualitative evaluations, such as ERO's personal in-school reviews.

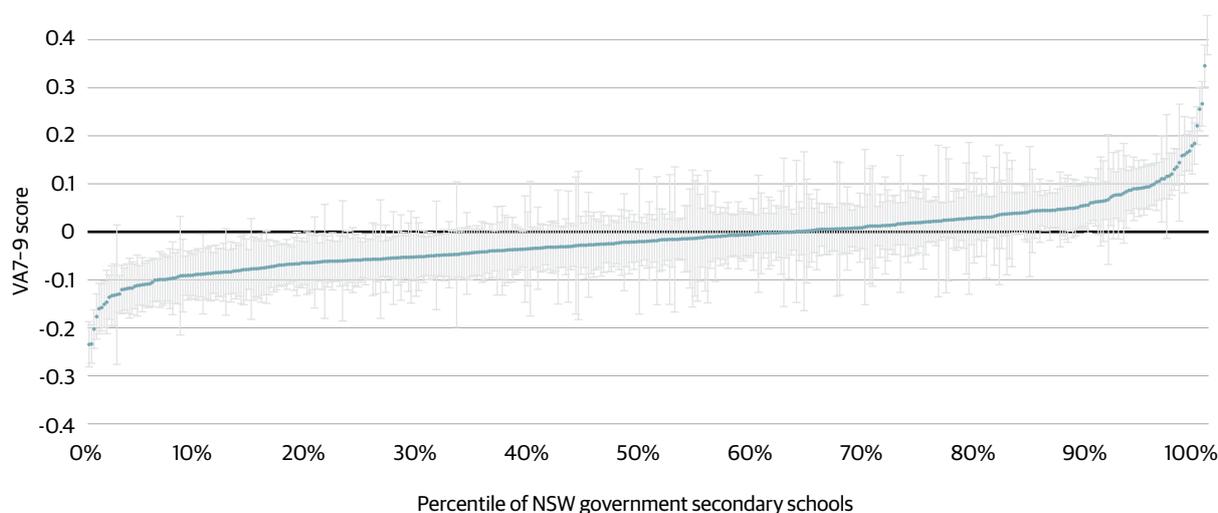
This is also typically done overseas when value-added models are used to evaluate school performance. In some countries, the value-added component only contributes 35% to the overall evaluation process.⁸¹ When combined, the evaluation method works symbiotically; the strengths of one method help overcome the weaknesses of another.

Most importantly, while we can see which schools seem to be doing very well and very poorly, we cannot tell what is going on inside the walls of the school. ERO's assessment could be broadened to look for school characteristics common to schools of varying performance outcomes.

Annual reports

Current ERO reviews only take place every one to five years, depending on the last review rating

Figure 10: NSW: School performance relative to all other government schools



Source: Centre for Education Statistics and Evaluation, "Using Value-Added Measures to Identify School Contributions to Student Learning" (Sydney: 2014), Figure 4.

a school has received; this is one weakness of ERO reviews that schools have complained about recently.⁸² Schools want more frequent feedback from the Ministry of Education.

If combined, our school performance tool would be able to fill the intervening time gap left by ERO by providing annual reviews for each school based on the NCEA performance of their students. Schools would be able to receive annual reports showing their current performance relative to every other school in the country, in addition to their current performance relative to previous years. Figures 10 and 11 show two examples from two NSW value-added reports.

Figure 10 illustrates individual school performance across all the secondary schools in NSW; Figure 11 illustrates an example of school performance over time.

These annual reports could also give schools additional information on their performance with different cohorts of students. They may find that some schools do better with Māori students or with students for whom English is

a second or other language (ESOL); other schools may do better with students with learning disabilities. The annual reports could provide this information.

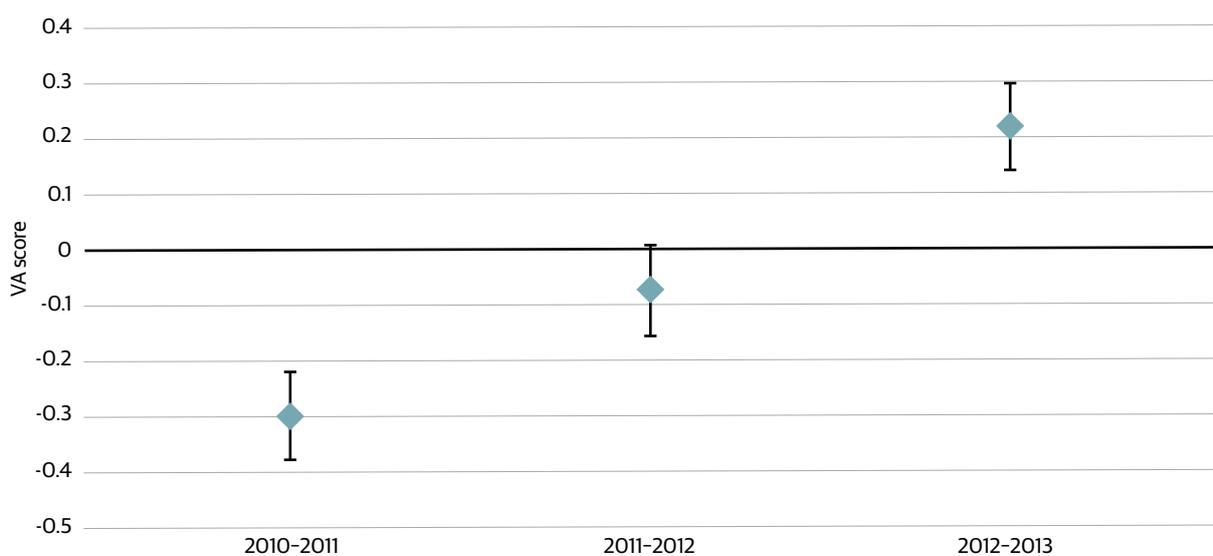
A future Initiative research note will present example annual reports from several New Zealand secondary schools that have kindly permitted us to use their school data in our project.⁸³

It will be up to the Ministry of Education to determine where and to whom these annual reports go: principals, boards of trustees, ERO, the public, or any combination of these.

Qualitative aspects and internal contribution

As described earlier, our tool's strength is in evaluating academic performance (our future reports will study later life outcomes, too). However, the tool cannot evaluate schools in qualitative areas such as school culture, music and sporting activities. Existing ERO reviews can fill this gap left by our tool. Personal in-school reviews

Figure 11: NSW: School performance over time



Source: Lucy Lu and Karen Rickard, "Value Added Models for NSW Government Schools" (Sydney: Centre for Education Statistics and Evaluation, NSW Department of Education and Communities, 2014), Figure 7.

by ERO are in a better position to evaluate the qualitative aspect of school performance.

Furthermore, internal reviews provided by a school's principals and teachers can and should also contribute to this component of school evaluation. An internal contribution is important because one criticism of ERO reviews is they only take place over one day, which critics argue does not always reflect a school's true performance over time. For parents, open school days and word-of-mouth play a larger role in evaluating the qualitative aspects of a school.

ERO also evaluates each school's governance process, including human resourcing and accounting. Our tool did not take any of these components into account when evaluating each school. While human resources and accounting are not the most glamorous components of a school, they are an essential background component of a school.

Learning from the top performers

Once our school performance tool has evaluated each school, the overall evaluation process does not stop there. Schools identified as top performers can be learned from, while schools identified as underperformers can be given additional support. Again, ERO is in the best position to complete this second step in the evaluation process as it requires in-school investigation. Two reports from New South Wales present a perfect example of this.⁸⁴

In *Sustaining Success*, the Centre for Education Statistics and Evaluation (CESE) identified 36 high-performing NSW government schools and explored six key drivers of school improvement and how they were implemented. Blaise Joseph from The Centre for Independent Studies (CIS) identified six key themes among nine disadvantaged primary schools. Their results are shown in Table 6.

Evaluation of any government education policy

In addition to its potential role within ERO, the Initiative's school performance tool could be used to evaluate any government education policy affecting school academic outcomes.

Current debates in education, such as the implementation of modern learning environments in New Zealand classrooms, could be systematically evaluated using our tool. In conjunction with the previous section, each in-depth evaluation of these top-performing schools could provide evidence for the success or failure of modern learning environments.

Our tool may find that modern learning environments only work in high-decile schools because the students there have more human capital to leverage, so the modern learning environment enables them to perform better academically. Equally, modern learning environments might not work in low-decile schools because the students there do not have the same human capital that students in higher decile schools have – students in low-decile schools may be served better by more traditional learning environments.

We do not know yet, but our tool could provide the government with the information to determine how effective modern learning environments are as an education policy.

In a similar vein, as part of the 2017 general election, New Zealand discussed closing all 11 charter schools established in 2013. Across all the debates, there was little conversation about any robust evidence of the effectiveness of charter schools. Charter schools could have been highly effective, or they could have been minimally effective, or they could have just been average; our tool could have provided the necessary objective information.⁸⁵ In the end, charter schools died under the 2018 Labour government, not because

there was evidence of their poor performance, but because of ideological bias against them.

Evaluation framework

Across these policy applications, our tool works best as part of an evaluation framework. It can be used for more than just producing annual reports; it can also be used broadly across a wide range of evaluation processes.

For example, if the Ministry of Education wanted to determine the effectiveness of state versus state-integrated schools, it could do that with our school performance tool. Equally, the Ministry could do this for state versus private or private versus state-integrated schools.⁸⁶

Furthermore, following the controversy around Bali Haque’s *Tomorrow’s Schools* report, our school performance tool could be used to test the effectiveness of several recommendations

Table 6: Six key drivers of success from high value-add schools in NSW

Centre for Education Statistics and Evaluation	The Centre for Independent Studies
High expectations: Creating and reinforcing high expectations for students, both academically and behaviourally.	School discipline: Based on high expectations, a clear set of consistently applied classroom rules, and a centralised school behaviour policy.
Student engagement: Classroom learning that is relevant to students’ lives and uses technology and innovative programming to enhance student engagement.	Direct and explicit instruction: New content is explicitly taught in sequenced and structured lessons. Includes clear lesson objectives, immediate feedback, reviews of content from previous lessons, unambiguous language, frequent checking of student understanding, demonstration of the knowledge or skill to be learnt, and students practising skills with teacher guidance.
Effective teaching: Using data to identify and respond to individual student learning needs, curriculum differentiation, and explicit teaching.	Experienced and autonomous school leadership: Stable, long-term school leadership, and principal autonomy to select staff and control school budgets.
Whole-school goals: Staff working together and setting shared goals to achieve school-wide improvements in student performance.	Data-informed practice: Using data from teacher-written, NAPLAN, and PAT assessments to improve teaching, track student progress, and facilitate intervention for underachieving students.
Collaboration: Sharing resources and taking a collaborative approach to planning, programming and assessing throughout the school.	Teacher collaboration and professional learning: Collaboration among teachers and specialist support staff to cater to the often complex needs of disadvantaged students. Focus on teacher professional learning; involving peer observations, mentoring, and attending practical professional development activities which help refine literacy and numeracy instruction.
Professional learning: Professional learning that supports strategic school goals and is shared among staff so that learning is embedded across the school.	Comprehensive early reading instruction: Five necessary elements of reading instruction: Phonemic awareness, Phonics, Fluency, Vocabulary, and Comprehension.

Source: Natalie Johnston-Anderson, “Sustaining Success: A Case Study of Effective Practices in Fairfield High Value-Add Schools” (Sydney: Centre for Education Statistics and Evaluation, 2017); Blaise Joseph, “Overcoming the Odds: A Study of Australia’s Top-Performing Disadvantaged Schools” (Sydney: The Centre for Independent Studies, 2019).

Note: It is worth noting that the CIS study looked at schools from across Australia and did not use a value-added model to identify top-performing schools – it instead identified low-decile schools that had an above-average attainment. Identifying high-performing schools using this method could be done in New Zealand by looking at low-decile schools with higher-than-average NCEA results – however, this method does not adjust for the background characteristics of the students in these schools.

in his report. This is particularly so for the recommendation to replace regional Ministry of Education offices with 20 regional education hubs.⁸⁷ If the Ministry set up one pilot regional education hub, our school performance tool could evaluate the effectiveness of the hub by evaluating the performance of schools within the hub before and after the pilot programme.

Moreover, if the regional education hubs were put in place nationally, our school performance tool could be used within the hubs to support their role as centres for collaboration. Our tool could be used to determine which schools are succeeding and thus learned from, while schools identified as underperforming can be given additional support.

Opening the black box of education

Since completing the first round of analysis with our school performance tool, New Zealand secondary schools remain as black boxes. However, in the process of developing our tool, we discovered multiple findings that have several implications for education in New Zealand. The first is empirical evidence that “decile is *not* a proxy for school quality”. The second is that 80% of our schools perform similarly once family backgrounds have been adjusted for. The third is there are still top performers, outliers that have been able to buck the trend and provide education outcomes above and beyond what would be predicted by the family background characteristics of their students. Importantly, these top-performing schools exist in both high- and low-decile schools.

This is not the end of the story. In this report, we have also outlined how the Ministry of Education could open the education black box by using our tool – particularly, the various ways it could be implemented through the Ministry.

Following the release of this report, the Initiative will be publishing research notes demonstrating what else is possible with our school performance tool. One research note will show example annual reports for individual schools that have given the Initiative permission to release their data from the IDI. Another will also show how schools perform when evaluated on later life outcomes such as progression to employment, tertiary education, and benefit uptake.

In addition to these research notes, there are several areas of interesting work that can be done; however, we are limited by time and the number of analysts at the Initiative. Nevertheless, this work should be done. We uploaded the SQL and STATA code behind our school performance tool to the IDI wiki so that any IDI researcher can take up this work in the future. Future economics master’s students would be ideal candidates. The Appendix elaborates on this.

This is just the beginning. What we have shown in this report is the first run (or the first version) of our model. There are still improvements to be made, both to our model and crucially to the education outcomes of New Zealand’s current and future students. The Initiative has provided the tool; the Ministry must now use it to open the black box of education. The wellbeing of New Zealand’s children is at stake.

Conclusion

New Zealand needs fair, data-driven evaluation of its secondary schools. For far too long, principals, teachers and parents have been left in the dark by the Ministry of Education. Anecdotal evidence and unreliable proxies have become the standard for how we evaluate our schools. The absence of better information has fuelled misconceptions about our school system and exasperated socioeconomic segregation in New Zealand's schools.

Meanwhile, our international performance in reading, mathematics and science has been declining; at the same time, New Zealand's education system has become one of the most unequal in the world.⁸⁸ New Zealand cannot afford to stay in the dark any longer – the wellbeing of New Zealand's current and future students is at stake – particularly for those from the most disadvantaged communities.

The Initiative knows New Zealand can and should do better. Using the world-leading data in SNZ's IDI, we built a school performance tool that demonstrates what can be done in New Zealand. The Ministry of Education can use our tool to fairly compare every secondary school in the country.

After a year-long analysis of nearly 400,000 students across 10 years, we found that the differences in school performance typically seen in NCEA league tables largely reflect differences in the communities those schools serve, not large differences in school quality or effectiveness.

We also showed for the first time that “decile is *not* a proxy for school quality”, and that there are high-performing schools across all deciles. In particular, 42 decile 1 and 2 schools outperformed 75% of every secondary school in the country when evaluated on University Entrance.

Importantly, while we have demonstrated that most schools (approximately 80%) perform similarly, this is not a call to keep the status quo. There are still underperforming schools across all deciles, the consequences of which are felt by students in those schools.

Most importantly, this report is a call to action for the Ministry of Education to use our tool and build an evidence base on “what works” to improve the outcome for every student in New Zealand. To open the black box of education and shine a spotlight on New Zealand's top performers. New Zealand is privileged to have a world-leading database like the IDI. However, without using it, without fair, objective and data-driven evaluation, we will remain in the dark and we will be setting New Zealand's current and future children up for failure.

Policy Recommendations

1. Every principal and school board of every secondary school in New Zealand should receive annual reports from the Ministry of Education tracking their school's performance using insights gained from our school performance tool. Currently, this requires every principal to agree to their school's data to be used in our tool. The Ministry is best placed to coordinate consent and provide these reports. For this to happen, parents must demand it from their school boards. Additionally, the Minister of Education must direct Ministry resources into producing these reports. Future work from the Initiative will demonstrate examples of these reports for three New Zealand secondary schools.

2. SNZ believes it is constrained in terms of allowing individual schools to be identified in IDI research, even anonymously, citing *Statistics Act 1975*. This appears as rule 5.14.2 in SNZ's *Microdata Output Guide*, and prevents the identification of individual entities, including schools. However, it is unlikely that Parliament's intent was to suppress anonymous data, such as individual anonymous points in a scatterplot, from an individual school. The Minister of Education should request the Minister of Statistics to ask SNZ to reconsider its interpretation of the Act. Parliament is considering updating the Act, and this is the right opportunity to address this oversight.
3. The Ministry of Education should build on the work the Initiative has commenced; all of our work is available in open-source form in the SNZ's IDI data lab. We urge the Ministry to follow our practice in making all of its work open-source for other IDI researchers to review and build upon as well.
4. The Ministry should use our school performance tool as part of a framework to evaluate new education policies. For example, the Ministry could test the effects of school practices like student streaming or the effectiveness of modern learning environments on student outcomes.
5. ERO should investigate whether its school performance assessments lead to similar conclusions as those from our school performance tool. In other words, are the top-performing schools identified by ERO the same top-performing schools identified by our tool? ERO should also use our school performance tool to assess whether there are institutional differences across schools that lead to better education outcomes for students. See Table 6 as an example of this kind of work.
6. The Ministry of Education is developing better data on in-school practice and staffing through the digital learning and data platform Edsby under Te Rito – New Zealand's student management system. This data, along with Novopay staffing data, should be integrated into the IDI to more comprehensively evaluate school performance.

Appendix

Weighted Relative Performance Index (WRPI) score

Our WRPI metric considers how a student performed in a standard, relative to all other students who completed that standard.

To illustrate, a student might receive an Excellence in the ‘Perform a solo or duet dance’ standard. This standard is taken by 1,000 students in the country, and 800 receive an Excellence grade. On the other hand, 1,000 students take ‘Apply the algebra of complex numbers in solving problems’, and 350 receive an Excellence grade. This means the second standard is likely the more challenging of the two.

An existing ‘percentile’ student achievement ranking helps solve this problem by taking an average of students’ relative performance. A student earning an Excellence in the dance standard is in the top 80% of students in that class, while the algebra Excellence puts a student in the top 35%. Averaging that percentile score across attempted standards builds a good measure of relative student performance but can unduly penalise students who push themselves with more difficult courses.

Our WRPI performs a similar calculation without penalising students for attempting more challenging standards. Our index is then:

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

where $WRPI_j$ gives the WRPI index score for student j ; α_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 .

$$x_{i,j} = \frac{\text{(number of students who sat standard } i\text{)}}{\text{(number of students who received the same or better grade than student } j \text{ on standard } i\text{)}}$$

Weighted NCEA score

Our weighted NCEA score assigns point values to different grades for each achievement and unit standard that a student sat. However, instead of applying arbitrary weights to different NCEA grades, we have used the empirical findings from Kamakshi Singh and Tim Maloney’s paper, which uses student NCEA data to predict university success.⁸⁹

Compared to the more uniform weights applied in the Cumulative Score, our weighted score puts more value on *merit* credits relative to *excellence* and *achieved* credits – this is based on Singh and Maloney’s findings that *merit* credits were more predictive of university success. Table 7 displays our relative grade weights compared to the Cumulative Score.

Table 7: Weights applied to NCEA grades

Weighted NCEA Score	Cumulative Score
Excellence: 4	Excellence: 4
Merit: 3.7	Merit: 3
Achieved: 1.36	Achieved: 2
Not achieved: 0	Not achieved: 0

Source: Kamakshi Singh and Tim Maloney, “Using Validated Measures of High School Academic Achievement to Predict University Success,” *New Zealand Economic Papers* 53:1 (2019), 89–106; Eric Crampton and Martine Udahehuka, “Score! Transforming NCEA Data” (Wellington: The New Zealand Initiative, 2018).

Future research notes

As noted at the end of Chapter 4 and throughout this report generally, the work presented here is only a sample of what our school performance tool can do. Future research briefly described earlier is elaborated below.

1. In the first upcoming research note, the Initiative will present example school reports from three New Zealand secondary schools that have kindly permitted us to use their school data in our project. This report will show how these schools have performed relative to every other secondary school in the country, in addition to how they have performed over time relative to themselves. These example reports are samples of what the Ministry could and should provide every school in the country. The fair, objective and empirical evidence presented in these reports would be invaluable for principals, boards of trustees, and education professionals reviewing schools through ERO.

For principals, particularly those from low-decile schools who do a great job with students from the most disadvantaged communities, these reports would highlight their stellar performance even in cases where their absolute performance in NCEA is below the national average.

These reports would help boards of trustees keep their principals in check when they do a poor job and conversely give praise when they do a great job.

For education professionals reviewing schools through ERO, information on a school's performance over time would be invaluable for analysing the effectiveness of different school interventions. In 2015, 67 school boards were under Ministry of Education intervention – for all the secondary schools in this sample, these reports would be able to indicate whether the current intervention should be dropped, boosted or tested further if the evidence of improvement were only moderate.⁹⁰

2. In a second research note, we will reveal how New Zealand secondary schools perform when evaluated on later life outcomes. This includes progression and completion of tertiary education, NEET (not in employment education or training) status, benefit uptake, and employment one, three and five years after college.

The purpose of this report will be to gain insights into whether schools that perform well on measures of NCEA also perform well based on later life outcomes. We may find that some schools performing well in NCEA underperform when evaluated on later life outcomes because they encourage students into easier NCEA standards rather than standards that prepare them for life after school. Alternatively, we may find that schools performing well in NCEA also perform well in preparing their students for later life.

3. Another research note will include insights on whether some schools perform better with different cohorts of students, in particular, Māori, Pasifika and students from disadvantaged backgrounds. We may find that some schools do better with disadvantaged students than others; we may also find that some schools do a better job with all students. It is possible that some students are traveling to a high-decile school across town when, in fact, the low-decile school down the road might be better suited to their needs. We do not know right now, but our tool could tell us. We may also find that the top 25 schools in the country are the best at teaching all cohorts of students. In that case, how can other schools learn from these 25 schools, or could those schools expand and grow to meet the demand for high-quality schools?

This area of research will contribute to ongoing debates in New Zealand on the performance of Māori medium schools, or private versus state schools versus state-integrated schools, or single-sex versus co-ed schools.

4. Finally, another research note will use data from our tool to generate example reports on potential post-school outcomes for different students presented by guidance counsellors in every secondary school in New Zealand. These reports would be aimed at students of varying backgrounds and highlight the successes of similar students in the past. For example, students with X background had great success with going into university while students with Y background had great success going into vocational training. For too long, the Initiative and education professionals have been worried that too many students are being pushed into university when they may be better suited for vocational training. These reports will help address that.
2. Many of the family background covariates (control variables) used in our project are the simplest version of that variable because of the 80/20 rule we applied to our school performance tool – our project is very much a proof of concept. Future master's theses could significantly improve upon the covariates we have used. For example, we have only used dummy variables for mother's and father's prison history; in other words, students were allocated values of 'yes' (1) or 'no' (0) based on their parent's presence in the corrections database. It would be more insightful for a future version of our model to include parent's length of stay in prison in addition to severity of offence. There are more than 40 covariates in our model – all of which could be refined and improved upon.

Interesting master's theses

Typically, the laborious and time-consuming task of data matching students with parents, and the high risk of insignificant results, inhibits one-year master's students from undertaking this type of work. However, we have already done all the data matching so all a student needs to do is run the Initiative's IDI code and start building on it. Example master's theses include:

1. Investigating the probability of enrolling into a high-performing school based on the location of a student's residence using meshblock data. As discussed in detail in the Introduction, many students attend high-decile schools out-of-zone. What is the probability that a student would be better served by his or her local low-decile school versus the out-of-zone high-decile school? This is a quantitatively difficult project and beyond the scope of our current research programme, but it would be an interesting research question worth investigating by a master's student.
3. Finally, because there is no standardised national testing performed and results collected in primary and intermediate schools, it is impossible to evaluate feeder schools using our school performance tool. This is important as many secondary schools often argue that they make up for academic weakness caused by feeder primary and intermediate schools. Using linked primary and intermediate school data in the IDI, a master's student could study feeder school quality using a modified version of our school performance tool. Again, this project is quantitatively difficult and beyond the scope of our current research programme, but it would be an interesting project for a master's student.

Endnotes

1. Harvey Goldstein, “Methods in School Effectiveness Research,” *School Effectiveness and School Improvement* 8:4 (1997), 369–395; Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools” (Sydney: Centre for Education Statistics and Evaluation, NSW Department of Education and Communities, 2014).
2. Thomas Dee and James Wyckoff, “Incentives, Selection, and Teacher Performance: Evidence from IMPACT,” NBER Working Paper 19529 (2013).
3. Students also have the option of completing the Cambridge International General Certificate of Education or an International Baccalaureate qualification during college. However, these options are not available in every secondary school in New Zealand.

Students typically gain NCEA level 1, 2 and 3 qualifications in years 11, 12 and 13, respectively – the three senior years in New Zealand secondary schools.
4. ERO disproportionately categorises low-decile schools in the one- to two-year review cycle (ERO’s poor performing category). ERO evaluates schools on various aspects of performance; the reason(s) low-decile schools are disproportionately represented in the poor-performing category is beyond the scope of this report. Tomorrow’s Schools Independent Taskforce, “Our Schooling Futures: Stronger Together” (Wellington: Ministry of Education, 2018); Education Review Office, “Our approach to school evaluations,” Website (17 July 2019).

However, evidence from a previous Initiative report, “Tomorrow’s Schools: Data and Evidence,” found no difference between low- and high-decile schools when evaluated on academic performance, after separating the effect of family background from average school performance across deciles. This report discusses this point further in Chapter 3. Joel Hernandez, “Tomorrow’s Schools: Data and Evidence” (Wellington: The New Zealand Initiative, 2019), 3.
5. For example, out of the 220,000 students who attended school in Auckland in 2018, approximately 127,000 attended a school within-zone, 38,000 attended a school out-of-zone, and 57,000 attended a school without an enrolment scheme. John Gerritsen, “Auckland schools overcrowded with out-of-zone students,” Radio NZ (23 April 2019).
6. HoonHo Kim and Diane Lalancette, “Literature Review on the Value-Added Measurement in Higher Education” (Paris: OECD Publishing, 2013).
7. One report used the NSW VAM to identify high-performing schools as ways to identify best practice. Natalie Johnston-Anderson, “Sustaining Success: A Case Study of Effective Practices in Fairfield High Value-Add Schools” (Sydney: Centre for Education Statistics and Evaluation, 2017).
8. Susan Edmunds, “Data shows NZ school zoning makes huge difference to house prices,” *Stuff* (4 August 2017).
9. Andrew Devonport, “The impact of secondary school enrolment schemes on school desirability, academic achievement and transport,” Thesis in fulfilment of Master of Geographic Information Science, University of Canterbury (2017).
10. Ibid.
11. Samuel Marsden Collegiate School, “Fees,” Website.
12. Adele Redmond, “Parents’ choice driving ‘eye-opening’ segregation in New Zealand schools,” *Stuff* (12 May 2017). We acknowledge that decile drift and the corresponding socioeconomic segregation are also influenced by personal biases and prejudices about certain cultures.
13. Tomorrow’s Schools Independent Taskforce, “Our Schooling Futures: Stronger Together,” op. cit. The *Stuff* school report even states on its website: “On their own, academic results do not indicate the quality of any school” and that “deciles have nothing to do with the quality of schools”. *Stuff*; “School Report: About,” Website.
14. Tomorrow’s Schools Independent Taskforce, “Our Schooling Futures: Stronger Together,” op. cit. 59. It is well established in the education sector that decile drift is occurring in New Zealand schools; however, part of the statistics cited here may be independent of decile drift and a result of changes in the distribution of school rolls over time.
15. Tomorrow’s Schools Independent Taskforce, “Our Schooling Futures: Stronger Together,” op. cit. 71.
16. Ibid. 69
17. In our final student-parent dataset, we evaluated 480 secondary schools. This figure is not an official statistic but a number produced after

- randomly rounding the school count to base 3 (RR3) in compliance with Statistics New Zealand’s confidentiality rules and regulations. The official number of secondary and composite schools that teach years 11–13 in New Zealand is 517; however, some schools in our dataset were dropped because they had fewer than 30 students. Education Counts, “Number of schools,” Website.
18. Tomorrow’s Schools Independent Taskforce, “Our Schooling Futures: Stronger Together,” op. cit.
 19. Social Investment Agency, “SIA’s Beginners’ Guide to the IDI” (Wellington: New Zealand Government, 2017).
 20. Statistics New Zealand, “Integrated Data Infrastructure,” Website.
 21. We removed (dropped) a small fraction of students from the dataset to meet certain criteria required for the evaluation model. A full list of the criteria can be found in the Appendix of our technical report: Joel Hernandez, “Separating School and Family: Evaluating the Effects of School and Family Background on Student Performance in NCEA” (Wellington: The New Zealand Initiative, 2019).
 22. A unit or achievement standard is the form of assessment used in NCEA.
 23. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit.
 24. Yugo Nakamura, “A primer on value-added models: Towards a better understanding of the quantitative analysis of student achievement,” Dissertation in fulfilment for Doctor of Philosophy, University of Washington (2013); Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit. 4. While it is well known in both the economics and education literature that family background influences student achievement, the causal mechanisms on how they affect student achievement is less understood.
 25. Dale Ballou, William Sanders and Paul Wright, “Controlling for Student Background in Value-Added Assessment of Teachers,” *Journal of Educational and Behavioral Statistics* 29:1 (2004), 37–65; Brian Rowan, Richard Correnti and Robert J. Miller, “What Large-Scale, Survey Research Tells Us About Teacher Effects on Student Achievement: Insights From the Prospectus Study of Elementary Schools” (Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, 2002); Daniel Muijs and David Reynolds, *Effective Teaching: Evidence and Practice* (London: Paul Chapman Publishing, 2001); and Kenneth J. Rowe, “The importance of teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision – Implications of findings from the international and Australian evidence-based research,” Supporting paper to invited address presented at the Making Schools Better conference: A Summit Conference on the Performance, Management and Funding of Australian Schools (2004).
 26. See the average percentage of variance explained in Table 4.2 in Robert J. Marzano, *A New Era of School Reform: Going Where the Research Takes Us* (Auroa, Colorado: Mid-continent Research for Education and Learning, 2000), 46.
 27. Value-added models also measure progress between two or more years – the exact length of time varies across countries, states, districts and cities. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit.
 28. Yugo Nakamura, “A primer on value-added models: Towards a better understanding of the quantitative analysis of student achievement,” op. cit.
 29. HoonHo Kim and Diane Lalancette, “Literature Review on the Value-Added Measurement in Higher Education,” op. cit. 4.
 30. Ibid. 4.
 31. Ibid. 5.
 32. William Sanders and Sandra Horn, “Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research,” *Journal of Personnel Evaluation in Education* 12:3 (1998), 247–256, 1.
 33. Thomas Dee and James Wyckoff, “Incentives, Selection, and Teacher Performance,” op. cit.
 34. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit. 7.
 35. Thomas Dee and James Wyckoff, “Incentives, Selection, and Teacher Performance,” op. cit.
 36. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit. 7.
 37. SCORE, “Measuring Student Growth in Tennessee: Understanding TVAAS” (Nashville: 2014), 7.
 38. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit. 7–9.
 39. See the discussion on Figure 1, which shows that higher socioeconomic schools tend to have higher growth, relative to their students’ starting points. This indicates that the effect of socioeconomic background on student growth is cumulative, and it is not fully

- accounted for in students' prior scores. Lucy Lu and Karen Rickard, "Value Added Models for NSW Government Schools," op. cit. 7–9.
40. Ibid.
 41. My School, "Model Farms High School, Baulkham Hills, NSW, Student gain," Website; Lucy Lu and Karen Rickard, "Value Added Models for NSW Government Schools," op. cit. 7.
 42. SAS, "Technical Documentation of 2017 TVAAS Analysis" (2017).
 43. My School, "Home," Website.
 44. William Sanders and Sandra Horn, "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database," op. cit. 1.
 45. Lucy Lu and Karen Rickard, "Value Added Models for NSW Government Schools," op. cit.
 46. Edward W. Wiley, "A Practitioner's Guide to Value Added Assessment" (Boulder: University of Colorado, 2006), 25–27.
 47. In a two-step model, the first step is at the student-level and the second is at the school-level.
 48. Multilevel models can identify schools that have high value-added scores as a result of high value-add (good) teachers rather than a result of any school-specific factor. However, we recognise that in many cases, high value-add teachers are likely endogenous with high value-add schools.
 49. Lucy Lu and Karen Rickard, "Value Added Models for NSW Government Schools," op. cit. Not all multilevel models estimate teacher value-added scores – for example, the NSW multilevel model only estimates student and school value-added scores in a two-step model.
 50. George Leckie, "England's Multilevel Model Based Value-Added School League Tables: Measuring and Communicating Statistical Uncertainty to Parents," *Bulletin of the International Statistical Institute* (2013).
 51. Education Bureau, "Schools Value-added Information System Technical Manual" (Hong Kong: 2015).
 52. Eric Isenberg and Heinrich Hock, "Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010–2011 School Year" (Washington, DC: Mathematica Policy Research, 2011); Naihobe Gonzalez, Matthew Johnson and Brian Gill, "Value-Added Models for the Pittsburgh Public Schools, 2014–15 School Year" (Oakland, California: Mathematica Policy Research, 2016).
 53. Edward W. Wiley, "A Practitioner's Guide to Value Added Assessment," op. cit.
 54. The University of Canterbury's Centre for Assessment and Monitoring offers various value-added reporting services to New Zealand schools for a fee. The evaluation services are available for primary schools (PIPS), Intermediate schools (MidYIS) and secondary schools (YELLIS, VA-12 and VA-13). University of Canterbury, "Assessment and evaluation for schools," Website.
 55. Adele Redmond, "Research project aims to determine what effect schools have on students' success – if any at all," *Stuff* (14 November 2018).
 56. Objective, data-driven evaluation of schools, as described in this report, could have already been done by the Ministry of Education. However, politics and ideological bias in education have prevented this kind of work from being done in the past. Because this work has not been done so far, the Initiative is demonstrating to the public that it can be done, only at a significantly higher time cost given the fraction of resources the Initiative has compared to the Ministry of Education.
 57. See endnote 17.
 58. See endnote 3.
 59. Briar Lipson, "Spoiled by Choice: How NCEA Hampers Education and What it Needs to Succeed" (Wellington: The New Zealand Initiative, 2017).
 60. Two previous Initiative reports go into this issue in further detail: Briar Lipson, "Spoiled by Choice: How NCEA Hampers Education and What it Needs to Succeed," Ibid. and Eric Crampton and Martine Udahemuka, "Score! Transforming NCEA Data" (Wellington: The New Zealand Initiative, 2018).
 61. One key strength of our model is it does not rely on the common value-added assumption that prior performance captures the full effect of a student's socioeconomic background and that student socioeconomic background has no future effect on academic performance.

One key limitation of our model is that in the literature, prior performance (prior ability) is predicted to be the single greatest driver of future performance. However, for the purposes of our model, we are trying to adjust for differences in family background, not differences in prior performance. Importantly, we believe we have a sufficient suite of family background characteristics so that prior performance will be highly correlated with the control variables we have used.
 62. This diagram assumes that school effects are independent of the effects of student characteristics. School effects could be moderated by student characteristics. For example, schools could do better

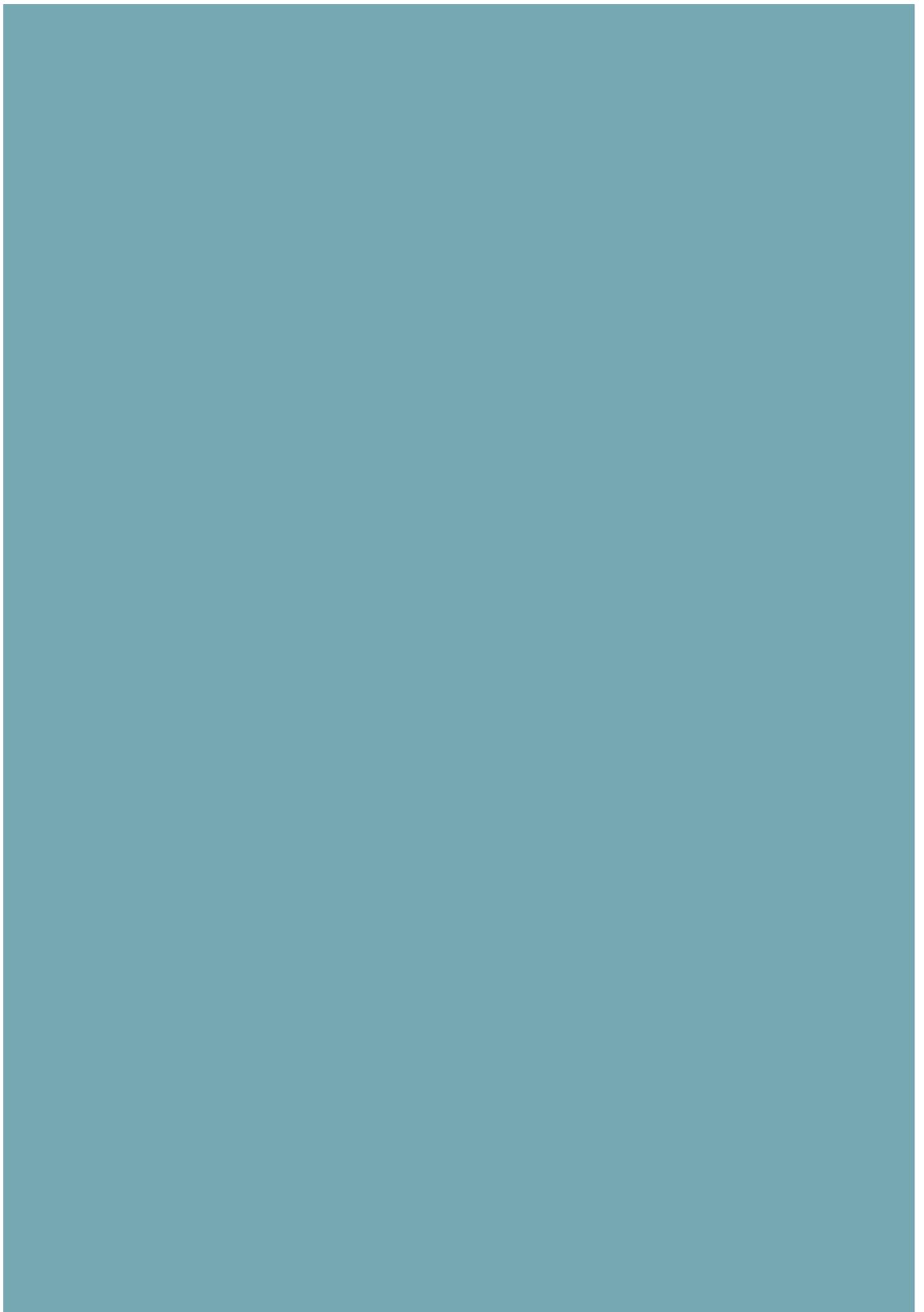
- with Māori, Pasifika, or Asian students for some set of unknown reasons. Because of this, future work from the Initiative will investigate the effect of school specialisation. On a technical note, this means we will be looking at the interaction effects between different ethnicities and schools. Further detail on future work is provided in the Appendix.
63. See endnote 17.
 64. The decision to use a fixed-effects model over a random-effects model is discussed further in the model background and specification section in our technical report. Briefly, however, a Hausman test indicates that the covariates included in our model are endogenous, and thus fixed-effects provides more efficient, consistent and unbiased estimates compared to random-effects.
 65. See endnote 17.
 66. In theory, it would be possible to link qualitative outcome variables such as overall life satisfaction from the New Zealand General Social Survey (NZGSS) to our model because the NZGSS survey data is available in the IDI. However, because the NZGSS does not collect data on every student, it is likely that there would not be enough data points to get reliable and useful results from this exercise.
 67. Variables 8 to 12 are all influenced by the school a student attends – as a result, the same student with the same behaviour may have different values for these variables depending on the school they attend. For this reason, it is recommended that these variables not be included in a future version of our model as they could introduce bias to our school estimates. It should be noted, however, that later robustness testing with models excluding these variables do not significantly affect our results. All robustness test results are presented in our technical report. Joel Hernandez, “Separating School and Family,” op. cit.
 68. Future iterations of our model should test the effects of school roll size as it is a common variable used in other value-added models implemented overseas. Additionally, future iterations of our models should also test differences between state, state-integrated and private schools as well as composite versus year 7 to 13 secondary versus year 9 to 13 secondary. The SNZ confidentiality rules did not allow us to compare state, state-integrated and private schools in this report, but future work could elaborate on our results.
 69. As noted in our “Tomorrow’s Schools” report, only the decile results for NCEA level 1 outcomes were presented. Joel Hernandez, “Tomorrow’s Schools: Data and Evidence,” op. cit. The complete results for NCEA level 2 and 3 outcomes are presented in the corresponding technical report. Joel Hernandez, “Separating School and Family,” op. cit. As mentioned in “Tomorrow’s Schools,” the decile results are almost identical for NCEA level 2 and 3 outcomes.
 70. Figure 4 is a duplicate of Figure 1 from Joel Hernandez, “Tomorrow’s Schools: Data and Evidence,” op. cit.
 71. Figure 4 also adjusted for time and school characteristics.
 72. If one were to use all the variables the decile system uses to calculate school deciles and use the same area-level data, one would expect to get a flat distribution of average school performance across deciles because we would be controlling for exactly what goes into calculating deciles. However, as discussed in Box 1, our model includes significantly more family background variables in addition to student-level data rather than area-level data compared to the decile system. For this reason, we did not expect to get a flat distribution of school performance across deciles.
 73. Note that our decile results in Figure 4 have a 95% confidence interval band, but the bands are so small they are not visible in the figure.
 74. Lucy Lu and Karen Rickard, “Value Added Models for NSW Government Schools,” op. cit.
 75. See Figure 4 in Centre for Education Statistics and Evaluation, “Using Value-Added Measures to Identify School Contributions to Student Learning” (Sydney: 2014). In Figure 4, NSW schools have remained anonymous in the figure showing the distribution of schools.
 76. Specifically, rule 5.14.2 in SNZ’s *Microdata Output Guide* prevents individual schools from being identified in any SNZ IDI research output, even in cases where a school’s results are anonymised. Statistics New Zealand, *Microdata Output Guide*, 4th edition (Wellington: 2016).
 77. A locally weighted scatterplot smoothing (LOWESS) curve is a fitted line applied to multiple data points in a graph. LOWESS curves are used when a linear line of best fit does not fit the data well.
 78. The red baseline indicates the performance of the one random secondary school on which every other secondary school is compared with. Importantly, value-added scores measure relative performance, not absolute performance. Changing the baseline school does not alter the results; it only shifts where the red baseline is located on the figure.
 79. Note that once the schools were tallied up by decile, the school counts were randomly rounded to base

- 3 and suppressed if the value was 5 or below in accordance with rule 5.1 (unweighted counts) and 5.9 (regression models) in SNZ's *Microdata Output Guide*. Statistics New Zealand, *Microdata Output Guide*, op. cit.
80. See endnotes 76 and 79.
 81. SCORE, "Measuring Student Growth in Tennessee: Understanding TVAAS," op. cit.
 82. Tomorrow's Schools Independent Taskforce, "Our Schooling Futures: Stronger Together," op. cit.
 83. Individual school results can be released from the SNZ datalab if the school's principal gives permission. Contact the author for more information.
 84. Natalie Johnston-Anderson, "Sustaining Success: A Case Study of Effective Practices in Fairfield High Value-Add Schools," op. cit. This is a follow-up paper to two earlier CESE publications: Deborah Bradford and Sophie Clarke, "High Value-add Schools: Key Drivers of School Improvement" (Sydney: Centre for Education Statistics and Evaluation, 2015), and Centre for Education Statistics and Evaluation, "Six Effective Practices in High Growth Schools" (Sydney: 2015). Blaise Joseph, "Overcoming the Odds: A Study of Australia's Top-Performing Disadvantaged Schools," op. cit.
 85. Our school performance tool could have provided information on charter schools that offered NCEA to their students. Not all charter schools offer NCEA, and thus not every charter school could be evaluated with our tool.
 86. Note that our school performance tool only evaluates schools based on NCEA results; therefore, some private schools that offer alternative qualifications such as Cambridge or International Baccalaureate to the majority of their students will not be evaluated effectively by our tool.
 87. The proposed regional education hubs would also take on some of the key responsibilities of boards of trustees such as hiring a principal. Tomorrow's Schools Independent Taskforce, "Our Schooling Futures: Stronger Together," op. cit.
 88. For a summary of New Zealand's declining performance in PISA and TIMSS, see the Ministry of Education, "PISA 2015 New Zealand Summary Report" (Wellington, New Zealand Government, 2016) and Briar Lipson, "Spoiled by Choice: How NCEA Hampers Education and What it Needs to Succeed," op. cit. For a description of the inequality in New Zealand's education system, see Yekaterina Chzhen, Gwyther Rees, Anna Gromada, Jose Cuesta and Zlata Bruckauf, "An Unfair Start: Inequality in Children's Education in Rich Countries" (Florence: UNICEF, 2018).
 89. Kamakshi Singh and Tim Maloney, "Using Validated Measures of High School Academic Achievement to Predict University Success," *New Zealand Economic Papers* 53:1 (2019), 89–106.
 90. Martine Udahemuka, "Signal Loss: What We Know About School Performance" (Wellington: The New Zealand Initiative, 2016), 6.

Bibliography

- Ballou, Dale, William Sanders and Paul Wright. "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational and Behavioral Statistics* 29:1 (2004), 37–65.
- Bradford, Deborah and Sophie Clarke. "High Value-add Schools: Key Drivers of School Improvement" (Sydney: Centre for Education Statistics and Evaluation, 2015).
- Centre for Education Statistics and Evaluation. "Six Effective Practices in High Growth Schools" (Sydney: 2015).
- . "Using Value-Added Measures to Identify School Contributions to Student Learning" (Sydney: 2014).
- Chzhen, Yekaterina, et al. "An Unfair Start: Inequality in Children's Education in Rich Countries" (Florence: UNICEF, 2018).
- Crampton, Eric and Martine Udahemuka. "Score! Transforming NCEA Data" (Wellington: The New Zealand Initiative, 2018).
- Dee, Thomas and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," NBER Working Paper 19529 (2013).
- Devonport, Andrew. "The impact of secondary school enrolment schemes on school desirability, academic achievement and transport," Thesis in fulfilment of Master of Geographic Information Science, University of Canterbury (2017).
- Edmunds, Susan. "Data shows NZ school zoning makes huge difference to house prices," *Stuff* (4 August 2017).
- Education Bureau. "Schools Value-added Information System Technical Manual" (Hong Kong: 2015).
- Education Counts. "Number of schools," Website.
- Education Review Office. "Our approach to school evaluations," Website (17 July 2019).
- Gerritsen, John. "Auckland schools overcrowded with out-of-zone students," Radio NZ (23 April 2019).
- Goldstein, Harvey. "Methods in School Effectiveness Research," *School Effectiveness and School Improvement* 8:4 (1997), 369–395.
- Gonzalez, Naihobe, Matthew Johnson and Brian Gill. "Value-Added Models for the Pittsburgh Public Schools, 2014–15 School Year" (Oakland, California: Mathematica Policy Research, 2016).
- Hernandez, Joel. "Separating School and Family: Evaluating the Effects of School and Family Background on Student Performance in NCEA" (Wellington: The New Zealand Initiative, 2019).
- . "Tomorrow's Schools: Data and Evidence" (Wellington: The New Zealand Initiative, 2019).
- Isenberg, Eric and Heinrich Hock. "Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010–2011 School Year" (Washington, DC: Mathematica Policy Research, 2011).
- Johnston-Anderson, Natalie. "Sustaining Success: A Case Study of Effective Practices in Fairfield High Value-Add Schools" (Sydney: Centre for Education Statistics and Evaluation, 2017).
- Joseph, Blaise. "Overcoming the Odds: A Study of Australia's Top-Performing Disadvantaged Schools" (Sydney: The Centre for Independent Studies, 2019).
- Kim, HoonHo and Diane Lalancette. "Literature Review on the Value-Added Measurement in Higher Education" (Paris: OECD Publishing, 2013).
- Leckie, George. "England's Multilevel Model Based Value-Added School League Tables: Measuring and Communicating Statistical Uncertainty to Parents," *Bulletin of the International Statistical Institute* (2013).
- Lipson, Briar. "Spoiled by Choice: How NCEA Hampers Education and What it Needs to Succeed" (Wellington: The New Zealand Initiative, 2017).
- Lu, Lucy and Karen Rickard. "Value Added Models for NSW Government Schools" (Sydney: Centre for Education Statistics and Evaluation, NSW Department of Education and Communities, 2014).
- Marzano, Robert J. *A New Era of School Reform: Going Where the Research Takes Us* (Auroa, Colorado: Mid-continent Research for Education and Learning, 2000).
- Ministry of Education. "PISA 2015 New Zealand Summary Report" (Wellington, New Zealand Government, 2016).
- . "School deciles: How deciles are calculated," Education.govt.nz.
- Muijs, Daniel and David Reynolds. *Effective Teaching: Evidence and Practice* (London: Paul Chapman Publishing, 2001).
- My School. "Home," Website.

- . “Model Farms High School, Baulkham Hills, NSW, Student gain,” Website.
- Nakamura, Yugo. “A primer on value-added models: Towards a better understanding of the quantitative analysis of student achievement,” Dissertation in fulfilment for Doctor of Philosophy, University of Washington (2013).
- Redmond, Adele. “Parents’ choice driving ‘eye-opening’ segregation in New Zealand schools,” *Stuff* (12 May 2017).
- . “Research project aims to determine what effect schools have on students’ success – if any at all,” *Stuff* (14 November 2018).
- Rowan, Brian, Richard Correnti and Robert J. Miller. “What Large-Scale, Survey Research Tells Us About Teacher Effects on Student Achievement: Insights From the Prospectus Study of Elementary Schools” (Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, 2002).
- Rowe, Kenneth J. “The importance of teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision – Implications of findings from the international and Australian evidence-based research,” Supporting paper to invited address presented at the Making Schools Better conference: A Summit Conference on the Performance, Management and Funding of Australian Schools (2004).
- Samuel Marsden Collegiate School. “Fees,” Website.
- Sanders, William and Sandra Horn. “Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research,” *Journal of Personnel Evaluation in Education* 12:3 (1998), 247–256.
- SAS. “Technical Documentation of 2017 TVAAS Analysis” (2017).
- SCORE. “Measuring Student Growth in Tennessee: Understanding TVAAS” (Nashville: 2014).
- Singh, Kamakshi and Tim Maloney. “Using Validated Measures of High School Academic Achievement to Predict University Success,” *New Zealand Economic Papers* 53:1 (2019), 89–106.
- Social Investment Agency. “SIA’s Beginners’ Guide to the IDI” (Wellington: New Zealand Government, 2017).
- Statistics New Zealand. “Integrated Data Infrastructure,” Website.
- . *Microdata Output Guide*, 4th edition (Wellington: 2016).
- Stuff. “School Report: About,” Website.
- Tomorrow’s Schools Independent Taskforce. “Our Schooling Futures: Stronger Together” (Wellington: Ministry of Education, 2018).
- Udahemuka, Martine. “Signal Loss: What We Know About School Performance” (Wellington: The New Zealand Initiative, 2016).
- University of Canterbury. “Assessment and evaluation for schools,” Website.
- Wiley, Edward W. “A Practitioner’s Guide to Value Added Assessment” (Boulder: University of Colorado, 2006).



School evaluation is an essential component of the education system. Identifying high-performing and underperforming schools is vital for building an evidence base for what works to improve education outcomes for students.

However, the task of evaluating schools objectively and without bias is extremely complicated.

The complex interactions between family socioeconomic background and academic performance result in measures of student NCEA achievement that are a mix of family, teacher and school effects.

A solution many countries have adopted is value-added models of assessment and evaluation.

Using linked administrative data in Statistics New Zealand's Integrated Data Infrastructure (IDI), the Initiative has built the country's first contextualised value-added model or simply, a school performance tool.

The Initiative's tool breaks new ground in New Zealand because it can identify how much each secondary school contributes to its students after separating the contribution of each student's family socioeconomic background.

This has allowed us to fairly and objectively compare low- and high-decile schools for the first time in New Zealand. In doing so, we found that 42 decile 1 and 2 schools outperform 75% of every other secondary school in the country when evaluated on University Entrance.

New Zealand has an opportunity to improve the outcomes for every student in the country by identifying top-performing schools using IDI data. This report outlines how our school performance tool can make education fairer in New Zealand.

\$25.00

ISBN

978-0-9951208-2-2 (print)

978-0-9951208-3-9 (online)



The New Zealand Initiative

PO Box 10147

Wellington 6143