

71517

The Development of the Student Assessment System in New Zealand

Lester Flockton



The Development of the Student Assessment System in New Zealand

Lester Flockton

© 2012 The International Bank for Reconstruction and Development / The World Bank
1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

1 2 3 4 15 14 13 12

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

Cover design: Patricia Hord.Graphik Design, Alexandria, VA

Contents

Abbreviations	v
About the Series	vii
About the Author.....	ix
Abstract	xi
Executive Summary	xiii
Introduction	1
National Large-Scale Assessment.....	2
Secondary School Examinations	5
Classroom Assessment.....	7
Drivers for Change.....	11
Lessons Learned	16
References	19

Abbreviations

AtoL	Assess to Learn
ERO	Education Review Office
GDP	Gross Domestic Product
NCEA	National Certificate of Educational Achievement
NEMP	National Education Monitoring Project
NZQA	New Zealand Qualifications Authority
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
PPP	Purchasing Power Parity
READ TF	Russia Education Aid for Development Trust Fund
SABER	Systems Approach for Better Education Results
TIMSS	Trends in International Mathematics and Science Study

About the Series

Building strong education systems that promote learning is fundamental to development and economic growth. Over the past few years, as developing countries have succeeded in building more classrooms, and getting millions more children into school, the education community has begun to actively embrace the vision of measurable learning for all children in school. However, learning depends not only on resources invested in the school system, but also on the quality of the policies and institutions that enable their use and on how well the policies are implemented.

In 2011, the World Bank Group launched Education Sector Strategy 2020: Learning for All, which outlines an agenda for achieving “Learning for All” in the developing world over the next decade. To support implementation of the strategy, the World Bank commenced a multi-year program to support countries in systematically examining and strengthening the performance of their education systems. This evidence-based initiative, called SABER (Systems Approach for Better Education Results), is building a toolkit of diagnostics for examining education systems and their component policy domains against global standards, best practices, and in comparison with the policies and practices of countries around the world. By leveraging this global knowledge, SABER fills a gap in the availability of data and evidence on what matters most to improve the quality of education and achievement of better results.

SABER-Student Assessment, one of the systems examined within the SABER program, has developed tools to analyze and benchmark student assessment policies and systems around the world, with the goal of promoting stronger assessment systems that contribute to improved education quality and learning for all. To help explore the state of knowledge in the area, the SABER-Student Assessment team invited leading academics, assessment experts, and practitioners from developing and industrialized countries to come together to discuss assessment issues relevant for improving education quality and learning outcomes. The papers and case studies on student assessment in this series are the result of those conversations and the underlying research. Prior to publication, all of the papers benefited from a rigorous review process, which included comments from World Bank staff, academics, development practitioners, and country assessment experts.

All SABER-Student Assessment papers in this series were made possible by support from the Russia Education Aid for Development Trust Fund (READ TF). READ TF is a collaboration between the Russian Federation and the World Bank that supports the improvement of student learning outcomes in low-income countries through the development of robust student assessment systems.

The SABER working paper series was produced under the general guidance of Elizabeth King, Education Director, and Robin Horn, Education Manager in the Human Development Network of the World Bank. The Student Assessment

papers in the series were produced under the technical leadership of Marguerite Clarke, Senior Education Specialist and SABER-Student Assessment Team Coordinator in the Human Development Network of the World Bank. Papers in this series represent the independent views of the authors.

About the Author

Lester Flockton is a graduate of Dunedin Teachers College and the University of Otago. He has extensive experience in New Zealand's school system as a teacher, principal, inspector of schools, Ministry of Education official, researcher, university teacher, educational thinker and leader. Throughout his career in education he has worked on many national curriculum committees and projects, including the revised New Zealand Curriculum (2007), led numerous professional development programs, made dozens of conference presentations, and held office in various professional organizations. He and his colleague Terry Crooks were the prime developers and co-directors of New Zealand's National Education Monitoring Project, which is widely recognized nationally and internationally. Lester maintains a wide-ranging interest and involvement in matters impacting on teaching and learning, curriculum and assessment, and the leadership, governance and management of schools. He examines the claims of research, policy and practice with a critical mind. Lester is a Senior Research Fellow and Emeritus Director of the Educational Assessment Research Unit at the University of Otago, and the National Education Monitoring Project. He has received a number of honours in recognition of his service to education in New Zealand.

Abstract

New Zealand is a relatively high-performing country according to international assessments. It has transited the journey to develop a strong student assessment system. This paper describes this journey, highlighting drivers that helped shaped the assessment system, and drawing lessons for other countries aiming to undertake this journey. This case study shows that New Zealand revamped its assessment system in the context of broader education and curricular reforms. A new vision of assessment for learning was realized in national large-scale assessment, secondary school examinations, and classroom assessment activities. Key drivers that allowed these reforms to take place were strong leadership at all levels of the education system, clear vision and agreed upon principles, stakeholders who were assessment literate, and regulations that gave discretion to the schools to self-monitor their performance.

Executive Summary

New Zealand has already transited the journey to develop an effective student assessment system. Countries aiming to develop their own assessment system may find in New Zealand's story important lessons for transiting this journey. Learning from this experience would put them in a better position to develop their own assessment system, encompassing national-large scale assessments for monitoring education quality, examinations for certification and selection of secondary school students, and classroom assessment for improving teaching and learning.

The development of the assessment system in New Zealand is closely related to the introduction of important reforms in the education system. First, reforms allowed for the creation of the right institutions for running and supervising the assessment. Second, a curricular reform shaped the content, skills, and formats of all assessment activities.

New Zealand's story shows that a good assessment system depends on a good curriculum, that the two are inseparable. Curriculum and assessment need to be seen as integral parts of a whole within contexts of teaching and learning.

New Zealand also introduced changes to the assessment system as a consequence of the expansion of the school system. This expansion pushed for the development of a secondary school examination program focused on providing more educational opportunities, rather than on selecting and removing students from secondary and tertiary education.

In New Zealand, the development of each type of assessment did not happen in parallel. The national large-scale assessment was commissioned by the Ministry of Education to a local university. This assessment ran on a regular basis for 15 years; however, in 2010, the contract for this program with the university was not renewed. As for 2012, a new assessment program is under study. Nowadays, New Zealand relies more on examinations and classroom assessment, and less on national large-scale assessments. International assessments continue to be used by government for comparisons of student achievement with other countries.

The secondary school examinations and classroom assessment activities are more stable features of the New Zealand school system. Examinations (Qualifications) have a long tradition in the country. In the 1990s, a new institution was created to administer qualifications and examinations, and a revised examinations framework was adopted based on the reformed curriculum.

Regarding classroom assessment, New Zealand has strong mechanisms in place to support and supervise the quality of classroom assessment activities. The curricular reform of the 1990s led to changes in scoring criteria for classroom assessments, and to a more decentralized approach to monitor school quality and classroom assessment activities.

A number of lessons can be drawn from New Zealand's story. First, strong leadership is required to put the system in motion and maintain its impetus. This leadership has to be present at all levels of the school system, and not only at the high political and governmental level. Second, leadership has to translate into a clear vision and agreed upon principles. Third, curriculum reform is part and parcel of assessment reform. The quality and worth of assessment is to a large measure determined by the quality and worth of curriculum and the alignment between the two. Fourth, a strong regulatory framework is important to ensure correct implementation; however, these regulations should give sufficient discretion to schools. Finally, stakeholders need to be assessment literate to fully benefit from an assessment system. Knowing how to interpret and use assessment information is critical to the development of an effective system.

The Development of the Student Assessment System in New Zealand

Lester Flockton

Introduction

Developing effective student assessment systems that can contribute to improve education quality is a priority for many countries. More and more, governments are interested in learning from the experience of other countries that have already made the journey (or are in the way) to develop their assessment system.

The purpose of this case study is to show the journey transited by New Zealand to build better student assessment system, and to offer lessons to other countries aiming to develop their own assessment system. This case study tracks the development of three types of assessments: national large-scale assessments (national monitoring standards), examinations (qualifications), and classroom assessments (school-based assessment). The paper offers a general description of each assessment type, showing both their current status and changes. It also presents the mechanisms that allowed for changes to take place, and analyzes the drivers that allowed for those changes to occur.¹

New Zealand is an interesting case study because it is a relatively high-performing country that has a relatively strong assessment system. In the international assessment of 2009 Program for International Student Assessment (PISA) (15-year-old students), New Zealand ranked 7 out of 65 countries. In other international assessments—namely the 2007 Trends in International Mathematics and Science Study (TIMSS) and the 2006 Progress in International Reading Literacy Study (PIRLS)—its performance has been nearer the international average.

New Zealand is a developed country that has a strong education system. It is a high-income country with a gross domestic product (GDP) at purchasing power parity (PPP) per capita of US\$28,250. It has a population of 4.4 million, three-fourths of which live in urban areas. It is an ethnically diverse country, with nearly 70 percent of the population being European descent, 15 percent Maori, and 9 percent Asian, among others. School coverage is universal from five years of age through secondary education. Over half of the youth population holds a tertiary education degree, and adult literacy rate is 99 percent.

¹ For a general discussion of the enabling context for student assessment, see Clarke (2012).

The New Zealand experience shows that a journey towards significant development and improvement across the interface of curriculum and assessment requires deep change in both policy and practice. Worthwhile change cannot happen overnight; it takes time, yet requires strategy, pace, and shared resolve at every level of leadership. They require the kind of leadership that is committed to seeing key principles borne out in practice, and a willingness among those in leadership positions to see their own ongoing learning as central to the process.

Moreover, if development and improvement are to be sustainable, they must be consistently underpinned by a clear vision. This vision should translate into a few strong principles that are understood by the widest possible community of interest or stakeholders: public, politicians, and professionals alike. Above all, worthwhile change requires ongoing investment in the widespread growth of practitioner capability at every level of the system.

This paper analyzes the assessment system in New Zealand, its current status, mechanisms of change, and drivers that allowed for those changes to take place. After this introduction, the paper presents the national large-scale assessment program, its main features, and mechanisms that were put in place for the program to operate. The following sections present similar information for the secondary school examinations and for classroom assessment activities. The fifth section reflects on the drivers that allowed for the development of the assessment system, highlighting the key role played by the curriculum. The last section synthesizes the main lessons for other countries aiming to develop their assessment system.

National Large-Scale Assessment

This section reviews the evolution of New Zealand's national large-scale assessment program. The first subsection presents a general description of the program, while the second subsection presents the mechanisms that allowed for the program to take place.

General Description

New Zealand had a national large-scale assessment program, the National Education Monitoring Project (NEMP),² which operated on a regular basis from 1995 until 2010. The purpose of NEMP was to collect information over time on how well students were reaching national learning standards stated in the national curriculum, and to inform where improvement might be needed (Ministry of Education 1993). The information provided would allow successes to be celebrated and priorities for curriculum change and teacher development to

² National Education Monitoring Project: <http://nemp.otago.ac.nz>.

be debated more effectively, with the goal of helping to improve the education that New Zealand's children receive.

NEMP assessed students at two target grades: year 4 (8–9 years old) and year 8 (12–13 years old). Year 4 was chosen because it is the first grade where most students have the skills to participate in rich assessment tasks. Year 8 is a transition point from primary to secondary schooling. New Zealand chose not to include a third grade in the final years of compulsory schooling for reasons to do with secondary school examinations. Nationally representative samples of students were selected to participate in each target grade. Near 3,000 students from around 120 schools participated in the assessment every year.³

The assessment program provided a detailed and comprehensive picture of trends in educational achievement nationally. Standardized procedures were used to ensure the validity and reliability of the assessment results. NEMP was administered once a year in the second half of the school year. Nationally, the exercise lasted five weeks. In each school, students had to work for about three to four hours spread over a period of five days.

NEMP covered the full breadth of the curriculum, rather than limiting to the so-called "basics." Different areas of the curriculum were tested in every year, with the same areas of the curriculum being tested again every three or four years. This ensured a stronger alignment between curriculum and assessment, recognizing and valuing the importance of a broad and balanced approach that integrates both aspects of the education process. Among the areas of the curriculum included in the assessment were reading, writing, mathematics, science, technology, social studies, physical education and health, art, and music.

NEMP used a variety of formats to assess student performance in the curriculum. Tasks range included paper and pencil tests, interviews, videos, performance-based tasks, small science experiments, reading from books, writing letters, dramatization with puppets, making art works, singing and dancing, and physical agility. These tasks had relevance to the lives of students (authenticity) and engaging appeal.

NEMP contributed to strengthen the teaching force in New Zealand. This contribution was the result of involving teachers in the assessment design, administration, and in the marking (scoring) of student tasks. After 15 years of NEMP, 1,365 teachers benefitted from valuable opportunities to extend their knowledge and skills in assessment as a result of this in-depth experience.

Each year the assessments were administered by about 100 experienced teachers seconded to the project for a period of six weeks. Following a week of training at a national workshop, the teachers worked in pairs for five weeks to conduct assessments in the selected schools. During that time, each teacher assessed 60 students. Teachers were not assigned to their own schools to administer the assessment.⁴

³ National Education Monitoring Project: http://nemp.otago.ac.nz/PDFs/key_features/chapter1_06B.pdf.

⁴ National Education Monitoring Project: <http://nemp.otago.ac.nz>.

Teachers also played an important role in carefully guided and supervised marking of student tasks. Marking using common criteria aligned with the curriculum requires high levels of professional judgment. This is a core competence that teachers can then transfer to the marking of classroom assessment tasks.

NEMP produced annual assessment reports informing how well student were reaching national learning standards over time. The reports gave detailed task-by-task results that show strengths and weakness in student performance. They also informed about differences in performance for different demographic groups and categories of schools (e.g. student gender, ethnicity, home language, and socio-economic level of school). The reports also informed about changes in student performance from one assessment round to the other. About two thirds of the tasks used in the assessment each year were made available for general classroom use.⁵

The NEMP results triggered a national discussion among educators, which was the base for pedagogical change. The widely circulated reports were accompanied by an initial response to the reports from a national forum of educators. The forum's comments helped begin debate about the meaning and implications of the reported results. This debate was a critical step for promoting changes in teaching and learning at the school level.

Mechanisms of Change

NEMP was originally commissioned by the Ministry of Education to a local university (Educational Assessment Research Unit, University of Otago). The commission stated that the university would be in charge of running all aspects of the assessments, including design, administration, and reporting, from 1995 to 2010.

The unit in charge of the assessment had a clear organizational structure. The project was led by the unit co-directors, 10 other staff, a national advisory committee, a Maori reference group, and curriculum advisory panels. Each year the project employed 96 practicing teachers on short-term contracts to administer the assessments in schools, and 180 practicing teachers and 45 senior education tertiary students to assist with the marking of the assessments.

Wide consultations took account of professional and community interests and each year schools, teachers, and students were involved with task development and trialing. Wide consultations took account of professional and community interests and each year schools, teachers, and students were involved with task development and trialing.

NEMP was funded by the Ministry of Education and had a yearly budget of about US\$2 million, less than 0.1 percent of the budget allocation for primary and secondary education. Almost half of the funding was used to pay for the time

⁵ National Education Monitoring Project: <http://nemp.otago.ac.nz>.

and expenses of the teachers, who assisted with the assessments as task developers, teacher administrators, or markers.⁶

The assessment operated on a yearly base between 1995 and 2010. After that, the Ministry of Education discontinued the program and began to replace it with one that focused largely on the assessment of reading, writing, and mathematics in line with the government's 2010 introduction of National Standards in these three areas. Beyond the contract with the university, the government did not have any obligation to ensure the continuity of the assessment program.

Secondary School Examinations

This section reviews the creation of the new secondary school examinations, the National Certificate of Educational Achievement (NCEA),⁷ which came to replace the previous School Certificate. The first subsection presents the historical context in which this change occurred, together with the new institutional arrangement put in place. The second subsection presents information on the enabling context and mechanisms of change that allowed for changes to take place.

General Description

For over 50 years students in the final three years of secondary school had been assessed with national examinations (qualifications) through end-of-year exams, moderated classroom assessments (school-based assessments), or a combination of both. The form of these assessments had been adjusted several times.

Through most of the 1950s and 1960s, with the school leaving age set at 15 years, the first major examination for students, School Certificate, was in Year 11 (students 15–16 years old). In most subjects, students were judged solely on a three-hour written national examination. To pass and be allowed to proceed to the next higher level in secondary school, students needed to average 50 percent or better in four academic subjects. The standards were set so that only about half of all candidates would pass.

Those who succeeded could then proceed to University Entrance examinations in their next year of schooling. This consisted of three-hour national examination papers in each subject and required 50 percent or better in each of four subjects to pass.

Many schools were permitted to “accredit” a quota of students in this qualification on the basis of school work and school examinations, the quota being based on the school's pass rates in School Certificate. Students who gained University Entrance were entitled to proceed to university study, or they could

⁶ National Education Monitoring Project: http://nemp.otago.ac.nz/PDFs/key_features/chapter1_06B.pdf.

⁷ NCEA Website: <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/>.

remain at school for a further year and sit University Bursaries examinations or University Scholarship examinations.

In the mid-1960s, about 60 percent of secondary students who remained at school after age 15 attempted School Certificate. About 20 percent of those who continued to remain at school attempted University Entrance. It is obvious, therefore, that a large percentage of students left secondary school with no formal qualifications. The system was designed to either reject or promote, with major implications for life opportunities.

By 1990, the percentage of students staying in secondary school to the senior years had increased dramatically. The employment market had tightened from the 1960s, so the government was keen for students to remain longer in the education system, yet many were not intending to go on to university.

As part of the education reforms, the New Zealand Qualifications Authority (NZQA)⁸ was established in 1990 with responsibility for qualifications awarded in secondary schools.⁹ It was charged with developing qualifications systems that provided varied pathways towards appropriate qualifications for as many students as possible, not just those who survived the percentage pass rates dictated by School Certificate and University Entrance.

NZQA adopted a single Qualifications Framework for both academic and vocational qualifications from senior secondary schooling through to tertiary qualifications. In 2002 the NZQA implemented NCEA as the main senior secondary school national qualification.

Over the years, NCEA has undergone a number of reviews and improvements. It challenges students of all abilities in all learning areas, and shows credits and grades for separate skills and knowledge. Using a range of assessment procedures, it enables students to gain credits from both traditional school curriculum areas and alternative programs. The qualification is recognized by employers and is used for selection by universities and polytechnics, both in New Zealand and overseas. It combines much of what is understood to be progressive educational and assessment practice in and for a modern world. Distinctive features are:

- Each year, students can study a number of courses or subjects, academic and vocational.
- In each subject, skills and knowledge are assessed against a number of standards.
- Schools use a range of internal and external assessments to measure how well students meet these standards.
- When a student achieves a standard, they gain a number of credits. Students must achieve a certain number of credits to gain an NCEA certificate.

⁸ NZQA Website: <http://www.nzqa.govt.nz/>.

⁹ NZQA is also responsible for the quality assurance of nonuniversity tertiary training providers.

- There are three levels of NCEA certificate, depending on the difficulty of the standards achieved. In general, students work through levels 1 to 3 in years 11 to 13 at school.
- Students are recognized for high achievement at each level by gaining NCEA with Merit or NCEA with Excellence.

Mechanisms of Change

The change from the School Certificate to NCEA required important adjustments in terms of policies, institutional structures, and human and fiscal resources. The most important of these changes was the creation of NZQA.

NZQA is backed by law and was created under the Education Act of 1989. It is a stable, independent government organization that reports to the Minister of Education and the Minister for Tertiary Education, Skills, and Employment. It is headed by a board appointed by the Minister of Education, and has a Chief Executive supported by four divisions: Qualifications, Quality Assurance, Office of Deputy Chief Executive Maori, and Strategic and Corporate Services. The Qualification division oversees all tasks related to secondary school examinations. The Chief Executive is also supported by external advisory groups.

In 2010, the staff employed at NZQA reached 401 people, almost all being on a full-time contract (NZQA 2010). A number of the staff had previously worked for the Department of Education and the School Certificate, before the creation of the NZQA. Additionally, new staff members were recruited after the creation of NZQA.

The NZQA is funded by both public and earned funding. In 2010, the total income reached almost US\$60 million, half of which came from public (Crown) funding.

Classroom Assessment

This section reviews the enabling context for classroom assessment activities in New Zealand. The first subsection analyzes the institutional arrangements in place to ensure the quality of classroom assessment activities. The second subsection reviews the mechanisms of change put in place to allow for the current level of development of classroom assessment activities in New Zealand.

General Description

The regulatory and curriculum reforms of the 1990s set a new frame for classroom assessment (school-based assessment) activities in New Zealand. Achievement objectives pertaining to eight incrementally stepped curriculum “levels” spanning primary and secondary schooling were to provide the reference for teacher judgments and reporting of achievement. The regulatory

stipulations (National Education Guidelines) were broadly stated, consistent with a self-management philosophy, giving schools considerable discretion over how they would manage their assessment processes.

However, the government audit agency (Education Review Office, ERO), which conducts on-site inspections of the work of every school every three or so years, had its own interpretation of what compliance with the guidelines should look like in practice. Because ERO's public reports on individual schools became high stakes, schools tended to bend their practices towards those that won the endorsement of ERO, despite widespread concerns among curriculum specialists, assessment experts, and others, and despite many school personnel themselves having negative attitudes to what they were being expected to do.

Classroom assessment became a mechanistic industry of check-listing large and unconnected numbers of achievement objectives in all subject areas for every student. The achievement objectives were variously specific or broadly stated, and their placement in one level rather than another was difficult to justify in terms of a demonstrably different or higher stage of learning and achievement. Summing up achievement (producing scores or grades for reporting) from large numbers of irregular objectives across multiple curriculum content strands was a flawed strategy because they did not lend themselves to credible aggregation.

Eventually the compelling arguments and representations of numerous critics of the system gained traction, particularly as a result of three government-requested reviews of the role and functioning of ERO. A major turning point occurred with the appointment of a new head of the ERO in 2002.

Accountability reviews, assurance audits, and effectiveness reviews, which were *done* to schools and gave little sense of ownership over the findings, were replaced with an evaluative approach that sought to assist schools to strengthen their own review strategies.

A new review framework introduced in 2002 recognized that schools themselves needed to be at the center of goal setting with robust processes for assessing and reviewing the effectiveness of their strategies for improving student achievement. They needed to learn how to get better at gathering, analyzing, and interpreting quality information to inform their judgments and direct their actions. Most important, they needed to *own* the processes and take responsibility for making them work. Mechanistic classroom assessment methods were no longer center stage.

The shift of direction by the ERO came together with the Education Standards Act of 2001 that would require schools to set goals, annual priorities, and targets for improvement of student achievement. In 2010 the government introduced regulations that required every primary school to set and measure student achievement against targets in the three areas of newly introduced National Standards: reading, writing, and mathematics. Each school would be required to monitor performance against these and report annually to the Ministry of Education and its community with an analysis of the differences

between the school's performance and "the relevant aims, objectives, directions, priorities or targets set out in the school's charter" (New Zealand Parliament 2001).

Consequently the variable and localized target setting and reporting from individual schools made it impossible to obtain a coherent national picture of student achievement in any one particular subject.

Because of concerns about significant achievement disparities between Maori and Pasifika and the rest of the student population, the analyses were to include specific information on the effectiveness of school strategies for those students. It was left to the individual school, however, to decide its curricular priorities for mandatory annual targets for improving student achievement.

Additional to the impact of new ERO approaches and the new planning and reporting legislation on school-based assessment practices was a major review of the national curriculum itself. While the new curriculum continues to include lists of achievement objectives organized into levels in each subject area, they are somewhat discretionary. The emphasis has shifted away from atomized bits of learning to a more coherent view of knowledge, concepts, and understandings, with integrated connections across subject areas.

Pedagogically, learning how to learn has become inseparable from the content of learning. Inquiry-based learning has emerged as a major curricular approach with significant implications for what is assessed, and how it is assessed. It is no longer commonplace to find schools check-listing against subject area achievement objectives. Rather, they are trending towards skills such as setting goals for inquiry; questioning; collecting, investigating, and analyzing information; creating something from the inquiry; sharing knowledge; and reflecting and self-assessing on the part of students.

Throughout the curricular and assessment reforms, considerable attention has been directed towards an overarching theoretical paradigm that separates assessment into two main areas of activity and function: assessment *for* learning, and assessment *of* learning, or formative and summative assessments. This paradigm, which emanated largely from assessment academics in England and the United States, was a reaction against testing driving teaching and learning, and testing dominating school-based practices. It won much support in New Zealand to the point that it became infused in Ministry of Education policy. With policy focus now squarely on learning, it was expected that classroom assessment practices should be totally consistent with this. The policy direction was made clear in the new national curriculum:

The primary purpose of assessment is to improve students' learning and teachers' teaching as both student and teacher respond to the information that it provides. With this in mind, schools need to consider how they will gather, analyze, and use

assessment information so that it is effective in meeting this purpose.

Assessment for the purpose of improving student learning is best understood as **an ongoing process** that arises out of the interaction between teaching and learning. It involves the focused and timely gathering, analysis, interpretation, and use of information that can provide evidence of student progress. Much of this evidence is “of the moment.” Analysis and interpretation often takes place in the mind of the teacher, who then uses the insights gained to shape their actions as they continue to work with their students

(Ministry of Education 2007)

Mechanisms of Change

In New Zealand, two government agencies play major roles in overseeing the work of schools: the Ministry of Education and the ERO. The Ministry of Education, working with the government, defines the resources available to schools, the curriculum to be followed, accountability systems, and the boundaries of acceptable practice within a self-managing system.

ERO is a public service department independent from the Ministry. It was established in 1988, in the context of a comprehensive reform to the education system. Its functions and budget are backed by law. ERO works as a quality assurance institution and is charged with evaluating, supporting, and reporting publicly on the performance of individual schools (ERO 2011). The ERO staff includes a chief executive, plus 150 review officers located in four regions.

The enabling context for classroom assessment also encompassed building teacher capability in assessment for learning. In 2002 the Ministry of Education established and funded a teacher development project titled “Assess to Learn” (AtoL), which is delivered by contractors in different parts of New Zealand. Schools that choose to do the program are required to commit to the Ministry’s outcomes:

- improved student learning and achievement
- shift in teachers’ knowledge and assessment practice
- development of coherence between assessment processes, practices and systems in classrooms and in schools so that they promote better learning
- a culture of continuous school improvement.

Commissioned research shows that the model, which is tailored to individual school’s specific goals, is proving successful in achieving all intended outcomes (Poskitt and Taylor 2008).

Since 2002 only about 35 percent of primary schools and 15 percent of secondary schools nationally have taken part in this program. Some have participated for one year; others for two or three years, which results in deep

learning and review of practice. Funding allocation, the availability of sufficient numbers of suitably qualified facilitators, and competing priorities for teacher professional development time are limiting factors, yet the program is having a valuable influence.

To support schools and teachers with analyzing, interpreting, and judging student achievement, the system has played a key role in supporting the development and use of a range of assessment tools that provide both norm and criterion referenced data. Efforts have been made to align assessment tasks and test items to curricular objectives, and in some cases, the curriculum levels.

In New Zealand there is no requirement on schools to use any particular tests or tools for classroom assessment; almost all schools selectively choose from those available from the Ministry of Education and the market. There are no mandated tests. It has become clear that unless teachers learn how to connect data to teaching and learning, the use of such tools becomes ritualistically mechanistic. Since teaching to tests is strongly discouraged in New Zealand, the value of tests for improving learning can be somewhat tenuous.

Drivers for Change

New Zealand's story of the development of national large-scale assessment, secondary school examinations, and classroom assessment has evolved over several decades. This evolution was possible thanks to a combination of factors that went far beyond the assessment system. This section reviews the main drivers that allowed for the reforms in the assessment system to occur, including reforms to the school system and a report calling to improve student assessment.

Reforms to the School System

A key factor to further developing the assessment system in New Zealand was an historical restructuring of the administration of the education system through a series of reforms that started in the early 1990s. In the school sector, these reforms were spelled out in *Tomorrow's Schools* (Ministry of Education 1988). The reforms encompassed a comprehensive review of curriculum and assessment.

The review of the administrative system was instigated by the Prime Minister, who was both Prime Minister and Minister of Education at the same time so that he could play a pivotal part in ensuring substantial change. The five-person task force he appointed to carry out the review and propose direction for the reforms included community, business, and professional people. The task force summed up the situation as it stood when it said, "Good people. Bad system" (Ministry of Education 1988).

The decision to address the administrative structure of the school system *before* tackling the core business of education itself (that is, teaching and learning, curriculum, and assessment and reporting) was critical. It was held that the right

sort of administrative environment was a prerequisite not only to enabling things to happen in different ways, but also in significantly better ways. Perhaps the most radical change to come out of the review, therefore, was the decentralization of school administration with a substantially different approach to regulation.

Schools became self-managing, regulation less prescriptive, and accountability more defined. The governance of each school became the responsibility of a Board of Trustees elected by parents and school staff. It was to be an enabling partnership between the community and the profession, with the principal having the status of a fully constitutional member of the Board, while also being responsible to the Board for the day-to-day management of the school. The New Zealand Department of Education was replaced by a Ministry of Education, with key responsibility for national policy and fiscal management.

Independent of the Ministry of Education, ERO was established to regularly audit and publicly report on the performance of individual schools in terms of the objectives in their charter “undertaking.” Every state-funded school was required by law to have a charter that sets out the goals, aims, priorities, and targets for the school consistent with government policy. The “undertaking” is a declaration of commitment by the school Board of Trustees to uphold and pursue those intentions.

The NZQA was also set up to oversee and increase the coherence of qualification systems so that there would be a unifying framework (The New Zealand Qualifications Framework) for all quality-assured qualifications in New Zealand, from secondary through to tertiary. Major reform to secondary school examinations and awards was also to be undertaken by NZQA. Traditional approaches were to be replaced by those that would better meet the needs of a diverse society and economy.

Several factors pushed for the reforms to the examinations. There was a perception that (i) the examinations were not fair to a large majority of the students; (ii) that instead of opening educational opportunities, the examinations were pushing students out of the education system; and (iii) that the examinations were not helping produce better job skills for a more competitive labor market.

Regarding classroom assessment, changes to the nature of regulation were particularly noteworthy. Prior to the reforms, all educational entities—from councils that governed tertiary institutions, to district education boards that controlled schools, to schools themselves—were required to operate within countless regulations authorized by the Minister of Education.

Consequently, the reforms led to the systematic dismantling of the regulatory system, replacing it with a new statutory approach called National Education Guidelines (Ministry of Education 1993). Schools were given responsibility for setting their own policies and practices within the scope and flexibility offered by the Guidelines; they were also required to substantiate the

effectiveness of their policies and practices. The curriculum and assessment section of the guidelines illustrate their open, nonprescriptive nature:

Boards of Trustees must foster student achievement by providing a balanced curriculum in accordance with the national curriculum statements (i.e. the New Zealand Curriculum Framework and other documents based upon it).

In order to provide a balanced programme, each Board, through the Principal and staff, will be required to:

- implement learning programmes based upon the underlying principles, essential learning areas and skills, and the national achievement objectives; and
- monitor student progress against the national achievement objectives; and
- analyse barriers to learning and achievement; and
- develop and implement strategies which address identified learning needs in order to overcome barriers to students' learning; and
- assess student achievement, maintain individual records and report on student progress.

(Ministry of Education 1993)

These Guidelines became the sum total of mandatory classroom assessment requirements. The majority of schools responded with renewed energy, ideas, and initiative. ERO nonetheless found that some struggled. In such cases ERO set expectations for improvement that it would check in a subsequent inspection. It was the responsibility of the school to seek out the kind of assistance it considered appropriate.

The curriculum and assessment reforms that followed two years into the administrative restructuring of the system were led by a new Minister of Education in a newly elected government.

The new government was committed to a curriculum restructuring that would enable stronger school accountability for student achievement. The sweeping changes it introduced, called the "Achievement Initiative," were largely influenced by curriculum and assessment reforms made in England in the late 1980s—reforms that proved to have little empirical support yet considerable ideological and political patronage. The rationale for these changes was improved student learning through a more structured national curriculum, more focused teaching and assessment, and greater continuity when students moved from one school to another. More insistent accountability through measurement and reporting were to give the rationale its leverage.

An overarching New Zealand Curriculum Framework (Ministry of Education 1993) introduced a new curriculum structure that shifted the focus

from content and approach to a policy based on outcomes. Each subject, or curriculum area, was organized into three or four major content strands, and within each strand a substantial number of achievement objectives were prescribed.

The objectives were organized into eight levels to represent the planned progress expected of students over their years of primary and secondary schooling. The achievement objectives and levels became the central focus for assessment and reporting. Showing the amount of progress against the levels was intended to raise accountability for the improvement of student achievement.

As intended, the national curriculum became the point of reference for assessment, reporting, and accountability. The model seemed logical and pragmatic. Yet almost from the outset it was under constant challenge from subject area experts, professional organizations, and practitioners alike.

Arguably the new curriculum system did offer teachers more structured guidance, particularly less experienced teachers and those not so confident in planning good programs without detailed expectations. And it did provide for some consistency from school to school. On the other hand, teachers and students fell into a treadmill of curriculum coverage and assessment “tick box” approaches that large lists of mandatory achievement objectives encourage. All of this typically leads to shallow and temporary learning rather than deep and sustainable learning. Moreover, associated assessment data were weak in validity and usefulness. Many teachers felt a loss of the trust, innovation, and exploration that they had previously enjoyed and valued in the working of the New Zealand education system.

Before the implementation of the new curriculum had been completed, the government had changed again. In 2000 the new government agreed to the Ministry of Education’s “stocktake” of the New Zealand Curriculum to investigate sector concerns and “the quality of these curricula in contributing to improved student outcomes” (Ministry of Education 2002).

Work on the revision, which followed a co-constructed development process (wide involvement of the education community), began in 2004. A draft version published in 2006 received overwhelming support nationally, and a confirmed revised national curriculum was launched by the Prime Minister in November, 2007. This new curriculum opened the way for improvements in the design of learning programs and the assessment of learning.

Report to Improve Student Assessment

A second significant event among many in New Zealand’s story of the development of assessment policy and practice was the report of a Ministerial Working Party titled *Assessment for Better Learning: Tomorrow’s Standards*, presented in 1990 (Ministry of Education 1990). The Working Party, comprising seven people from professional and academic fields, was set up at the request of

the Minister of Education to recommend to the government assessment procedures for (i) monitoring the effectiveness of the New Zealand school system on student learning, and (ii) assessing the impact of individual schools on student learning.

The Working Party on Assessment had to advise ways of reporting, taking account of different audience needs, and the possible effects of assessment and reporting procedures for students, teachers, the curriculum, schools, employers, and the wider community. Notably, there was consensus from public consultations used to inform the working party's report that assessment for judgmental and comparative purposes can be very damaging to students and schools. "But when its objective is to bring about improved learning and teaching, and it is seen as an integral part of the learning cycle, it can have very positive influences on the quality of learning" (Department of Education 1990). This sums up what has become foremost in the thinking behind the formulation of assessment policy and practice in the New Zealand system over recent years.

Regarding national large-scale assessment, a growing demand for information about education quality pushed for the development of New Zealand's NEMP. At a national level, government, policy makers, and the general public needed information on how well the school system was performing overall and how particular groups of students were achieving. Moreover, they were interested in knowing trends over time and the extent to which these might reflect policy initiatives and other changes. Up until 1995, there was no such information available in New Zealand, despite a number of governmental reports over a 50-year period strongly recommending that there should be national monitoring of standards.

It was the Ministerial Working Party on Assessment for Better Learning that brought such recommendations to a head. Their report put the case for national monitoring. It proposed a broad framework for how it could be administered, specifying grade levels to be tested, and stating that a wide range of curricula outcomes should be monitored.

Approval to proceed with the national large-scale assessment program rested with the Minister of Education. He was convinced that it was a necessary part of an overall assessment policy, and that it would be integral to the Curriculum Framework that he wanted in place. Thus he approved the national assessment to evaluate overall educational standards.

The Ministry called for registrations of interest to further develop and operationalize the national assessment design. A contract was awarded to the Educational Assessment Research Unit at the University of Otago, whose two co-directors proceeded to investigate, observe, and evaluate models of system monitoring in other countries, before proposing a detailed plan for New Zealand.

The critical preliminary step in the national assessment design required developing a vision based on solid principles that would ensure the integrity of

the assessment. Ten principles that would direct and underpin the development of the national assessment were identified:

1. Create nonthreatening assessment conditions.
2. Assess a broad range of achievements.
3. Choose only tasks that reveal important learning outcomes.
4. Include a wide range of task difficulty.
5. Use engaging task approaches
6. Present tasks in formats appropriate to the skills being assessed.
7. Involve practicing teachers throughout the process.
8. In the scoring/marking process, focus on what is important for learning.
9. Provide trustworthy information.
10. Contribute information for improvement.
11. These principles became the foundation of NEMP.

Lessons Learned

New Zealand's story reveals a number of key drivers behind the changes that were necessary for the goal of assessment for better information and learning to be achieved. These drivers tended to be common across policies and programs for different assessment components in the system, rather than specific to one or another. Paradoxically, many institutional and policy positions that were powerful drivers of change were in fact disabling of the kind of changes needed to achieve improvements of true and sustainable educational worth.

New Zealand's story shows that not all change is good or achieves the results intended. The substance and processes of change need to be thoroughly understood and their consequences carefully calculated. It is not enough to have strong leadership, whether political or professional; it needs to be leadership that is knowledgeable, insightful, well informed, properly analytic, and prepared to entertain ideas that may be at odds with its own. It is not enough to simply follow another system's model because ideology (political or professional), demographics, and fiscal capability seem to align with one's own. All models should be critically investigated for their merits and shortcomings with a view to selective adaptation. It is not enough to consider assessment as an entity in itself when developing policy and practice. Assessment is deeply entwined in contexts of curriculum, teaching and learning, and reporting, which are invariably underpinned by philosophic as well as pragmatic considerations.

Of the key mechanisms or drivers that have facilitated change towards improved assessment in New Zealand, five are arguably relevant to most systems:

- *Leadership and consensus building.* The New Zealand experience has shown that good leadership at every level (governmental to educators) is critical,

and that good leaders are highly active learners. Leadership effects are most beneficial for system development when political, administrative, professional, academic, and community representations work together; are respectful of others' expertise and ambitions; and are able to reach common ground and share common goals. When all such constituencies push together towards shared goals and how those goals should be operationalized, the chances of success are significantly brighter and the work required is more productive.

- *Clear vision and guiding principles.* The New Zealand experience has shown that first and foremost to securing sound assessment systems is the identification and statement of a clear vision and a set of principles that give clear ethical and quality boundaries for the development and working of the assessment system at every point. Principles express beliefs, values, and expectations that should underpin the goals, policies, and practices. They include the characteristics of quality assessment, and they are used regularly to review and check the integrity of intentions and actions. The process of developing principles in New Zealand is typically collaborative and consultative, with sectorwide opportunities for specialist and nonspecialist contribution and comment before the principles are formalized by those responsible for the process.
- *Regulation.* The New Zealand experience has shown that systemic improvement of assessment for information and learning typically requires some measure of regulatory reform to give opportunity for capability to be developed and expanded. Overprescriptive regulation tends to delimit the exploration needed for development, so there is a fine balance to be struck. New Zealand opted to reconceptualize regulation so that it sets a framework of expectations, yet gives sufficient discretion to schools over the form in which those expectations may be conducted.
- *Curriculum.* The New Zealand experience has shown that curriculum reform is part and parcel of assessment reform. Curriculum sets out the learning expectations for different levels of schooling; assessment is concerned with judgments about the degree to which the learning has occurred. Curriculum informs the scope of what should be assessed; assessment informs curriculum action (teaching and learning). The important implication is that the quality and worth of assessment is to a large measure determined by the quality and worth of curriculum.
- *Assessment literacy.* This might seem an obvious necessity for the improvement of assessment systems, but it cannot be overstated and should never be assumed. The New Zealand experience has shown that when there is a low level of assessment literacy among leaders and administrators themselves, the assessment system can suffer and become misdirected—at a cost. Teacher professional learning, however, is often the “make or break” of an assessment system, no matter how well

designed. New Zealand recognizes this, but funds needed for powerful and sustained learning across a system are considerable, and more than most governments can afford. Commitment to steady incremental improvement over time is the reality. All good change takes time, and improvement requires change to be understood as a continuous process rather than a one-off event.

The story of assessment in any education system must connect with the past in order to understand the present and construct possibilities for the future. Hopefully these lessons will help other countries build their own assessment story.

References

- Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." SABER–Student Assessment Working Paper No. 1. World Bank, Washington, DC.
- Department of Education. 1989. "Assessment for Better Learning: A Public Discussion Document." Wellington: Department of Education.
- Education Review Office (ERO). 2011. *Annual Report for the Year Ended 30 June 2011*. Wellington: ERO.
- Ministry of Education. 1988. *Administering for Excellence: Report of the Taskforce to Review Education Administration*. Wellington: Ministry of Education.
- — —. 1988. *Tomorrow's Schools. The Reform of Education Administration in New Zealand*. The Rt. Hon. David Lange, Minister of Education.
- Ministry of Education. 1990. *Tomorrow's Standards: The Report of the Ministerial Working Party on Assessment for Better Learning*. Wellington: Learning Media, Ministry of Education.
- — —. 1993. *National Education Guidelines*. Wellington: Ministry of Education.
- — —. 1993. *The New Zealand Curriculum Framework*. Wellington: Ministry of Education.
- — —. 2002. *Curriculum Stocktake Report to the Minister of Education*. Wellington: Ministry of Education.
- — —. 2007. *The New Zealand Curriculum*. Wellington: Learning Media for Ministry of Education.
- New Zealand Parliament. 2001. Education Standards Act. Wellington: New Zealand Parliament.
- New Zealand Qualifications Authority (NZQA). 2010. *Annual Report 2009–2010*. Wellington: New Zealand Qualifications Authority.
- Poskitt, J., and K. Taylor. 2008. *National Education Findings of Assess to Learn (AtoL) Report*. Wellington: Ministry of Education.

New Zealand is a relatively high-performing country according to international assessments. It has transited the journey to develop a strong student assessment system. This paper describes this journey, highlighting drivers that helped shaped the assessment system, and drawing lessons for other countries aiming to undertake this journey. This case study shows that New Zealand revamped its assessment system in the context of broader education and curricular reforms. A new vision of assessment for learning was realized in national large-scale assessment, secondary school examinations, and classroom assessment activities. Key drivers that allowed these reforms to take place were strong leadership at all levels of the education system, clear vision and agreed upon principles, stakeholders who are assessment literate, and regulations that gave discretion to the schools to self-monitor their performance.

Lester Flockton, *Emeritus Director, Educational Assessment Research Unit, University of Otago, New Zealand*