MONITORING SUBJECT VARIATION WITHIN A NEW ZEALAND STANDARDS-BASED ASSESSMENT SYSTEM

By

Dr T. M. Boustead
(therese.boustead@canterbury.ac.nz)

CEM Centre
College of Education
University of Canterbury
Christchurch
New Zealand

Paper presented at the 2008 IAEA Conference, Cambridge, UK

MONITORING SUBJECT VARIATION WITHIN A NEW ZEALAND STANDARDS-BASED ASSESSMENT SYSTEM

Abstract

In 2002, New Zealand moved from a subject-based external examination system to a standards-based generalized system with both internal and external examinations. There are now over 60,000 credited standards within the national framework in the senior secondary schools and beyond. These standards span a wide range of academic and vocational areas. Although one of the two categories of standards offers grades in achievement, merit or excellence, the focus of the system is on the number of standards passed, irrespective of the subject area or grade. The change to the recent system created a number of issues; among them an important one being the perceived loss of subject comparability. The New Zealand CEM Centre has been in a unique position to successfully monitor this issue using prior assessment from value added indicator systems as an anchor. The approach has highlighted comparative subject issues that can be hidden within a generalized standards-based system.

Introduction

Six years ago, New Zealand changed from a norm-referenced, subject-based, external examination system to a standards-based generalized assessment system with both internal and external assessments. New Zealand based its concepts for this new system on the Scottish Vocational Qualifications and applied the concepts to both vocational and academic areas at secondary and tertiary levels, except universities. The main difference in changing to a new standards-based system is that while students competed with each other in the norm-referenced system, students compete with standards in a standards-based system. The two systems are indicative of two main goals in Achievement Goal Theory, which is a framework for academic motivation (Dweck & Leggett, 1988; Nichols, 1984). Motivation through competition with others (performance approach) is similar to the old norm-referenced system, and motivation through own performance (mastery approach) is similar to the standards-based system.

However, in a recent student motivation survey, there were some perceptions that the standards-based system has inconsistent, unfair marking and grading criteria and practices; the required number of credits to be passed each year is perceived as a minimum rather than a maximum; the qualification design was not recognizing excellence and there was a demotivating outcome possible in some subject areas (Meyer, McClure, Walkey, McKenzie & Weir, 2008). This paper will focus on the senior secondary school assessment level, looking in particular at student motivation and subject comparability within the standards-based system.

The Old System – norm-referenced

Prior to 2002, the old norm-referenced system had a number of advantages. Subjects were clearly defined into traditional areas such as English, mathematics, economics, history and geography. Students received a clearly defined per cent score for each subject and, in Years 11 and 13, examinations were predominantly external and strictly controlled while Year 12 examinations were internally assessed. National results were monitored and adjusted to obtain marking parity between subjects. The motivating goal was for all students to pass the official national qualification with the more able students aiming to achieve as high a mark as possible.

However, there were disadvantages with the old system. Allocation of marks to internally assessed Year 12 results were based on the achievement of each school's students the year beforehand in School Certificate (Year 11) examinations. If one student improved, then by definition, another student regressed in that same school. Another major concern was the 50% pass rate. Students who scored below 50% were not given credit for any knowledge or skills they possessed.

The New System – standards-based

There are advantages to the new approach of a standards-based system. The largest advantage is the improvement in pass rate from 50% under the old system to about 80% under the new system. Credits are allocated to each standard and students are acknowledged for the credits they gain while working towards a total of 80 credits each year for the National Certificate in Educational Achievement (NCEA). The 80

credits are roughly equivalent to three subjects under the old system but, unlike the old system, the 80 credits can be obtained from a combination of standards from any number of curriculum areas. The only criterion is that eight literacy credits and eight numeracy credits must be included.

At Level 1 (Year 11) some standards are 'bottom-up', meaning that students could accumulate standards below Year 11 which contribute to NCEA Level 1. This is an advantage for students who usually struggle academically in school. Theoretically, the flexibility in choice of standards and a 'bottom-up' approach means that NCEA can be adjusted to an individual's need or a school's specialty. However, with the focus on the number of standards passed rather than a grade, it is possible for the record of learning for a less able student to contain more passed standards than a record of learning for a more able student. Possible explanations are in the next section. Difficulties also arise with the official notification of results ('record of learning') which can be confusing to parents and employers. "A" indicates a pass or achieved and an "E" indicates an achieved with excellence.

Although subjects are not clear in either the standards-based system or curriculum document (Ministry of Education, 2007), schools are still organized in terms of subject areas and students still think in terms of traditional subjects while perceiving differences between subject areas (Meyer, McClure, Walkey, McKenzie & Weir, 2008). While a generic approach to both curriculum and assessment is suitable for the junior secondary level and below, specialization is important at the senior secondary level for preparation into university and specialized careers. Within a standards-based generic system, the difficulty is to define what a subject is and to determine if there are any hidden patterns within and between the subjects that could impact on specialization.

Why Compare Subjects?

The lack of definite subjects within the standard-based system and the potential for overlapping subject areas has implications for the school, tertiary education and employment.

The standards-based system acknowledges the universality of learning skills and key competencies such as thinking, using language, symbols and texts, managing self, relating to others, participating and contributing across all curriculum areas (Ministry of Education, 2007). However, specialization of disciplines should be important for senior levels. Subject disciplines have their own perspective on the universe. For example, at first glance History and Archeology appear similar. However, their differences lie in the second-tier questions that they answer, the varied data they analyze and their different perspectives on the data (Black, 2000). Likewise, Mathematics involves a hierarchical accumulation of knowledge and skills, but the level of mathematics required for a specialization in senior secondary school is more complex and specialized than the mathematics level required for other subjects.

At the school level, a focus on generic skills and knowledge can lead to the development of important gaps in knowledge. For example, a student could gain NCEA Level 1 with 80 credits having registered for 20 Mathematics standards but only passing eight of those credits. This could result in a lack of preparation for higher levels, especially in a discipline that relies heavily on mastering prior knowledge and understanding. For the student, strong specialization at the senior level can also be important for preparation for tertiary study and some careers. For example, medicine requires a strong knowledge of mathematics and the sciences and both commerce and engineering require a university stage 1 level of mathematics.

Subject specialization at the senior level requires teacher specialization which has proven to have a positive impact on achievement. Research in the United States, shows that collective teacher quality (which includes content specialization) positively relates to student achievement, especially in reading and mathematics (Heck, 2007). The lack of clear specialized subjects and overlapping domains in the standards-based system can have implications for teacher qualifications, specialization and school restructuring of faculties and subject areas.

The subject focus of the revised senior secondary scholarships led to the first media-wide controversy regarding subject differences in the standards-based system. For example, in the first year of scholarships under NCEA there was a larger than expected variation between scholarships awarded to subject areas. A large number of

senior secondary scholarships were awarded to English but very few scholarships were awarded to the sciences, especially biology. With the lack of hard data about subjects under the new standards-based system, there was speculation that the more able students did English and those who chose the sciences were less able. The Curriculum Evaluation & Management (CEM) Centre, which is a unit focused on value added projects developed by the University of Durham CEM Centre (Fitz-Gibbon, 1997), was able to study this data from a different perspective. By tracing back to a common assessment of developed abilities, the CEM Centre found that both English and the sciences contained very able students. In fact, the top students in Year 13 sciences gained a higher weighted NCEA average than the top students in Year 13 English. Although scholarships were sat in addition to Year 13 NCEA Level 3 that year, there did not appear to be any evidence in the Year 13 weighted achievement results to support the variation between English and science scholarship allocation. The approach undertaken by the CEM Centre is described below using as an example the trace from a common baseline at Year 9 to subjects in a standard-based system at Year 11. The study focuses on consistent patterns.

Method: Accumulation of Standards to Form Subjects

The first difficulty in a generalized standards-based system is to define a subject. There are two types of standards. Internally and externally assessed Achievement Standards are more straightforward than internally assessed Unit Standards as they are allocated to a traditional academic domain and subfield, for example domain=physics, subfield=science. Approximately 24 credits are allocated to each academic subject domain. However, these standards can overlap several subject areas with their combinations of domains and subfields. Unit Standards, which are more vocationally orientated, could possibly be placed into an academic-related 'subject' solely by the nature of the assessment task.

Not all standards are accepted for entrance to university. The very nature of specialization in universities means that clarity of the generic-based system is needed for entrance into universities. There is now an official list of 'approved' subjects at Year 13 for entrance to university which can include standards from a range of

Achievement (AS) and Unit (US) Standards. Table 1 illustrates one approved Level 3 subject option.

Subject Discipline	*Standards from:*
Mathematics with Calculus	Domain Trigonometry Domain Geometry Domain Calculus AS90638, AS90639 US5267, US11102, US12344

Table 1: A Year 13 approved subject for entrance to university

To be able to compare subject-based results, the CEM Centre used a similar approach to that of the university entrance requirements but also included the grade (Achieved, Achieved with Merit or Achieved with Excellence). The CEM Centre defined subject areas based on the Achievement Standard domains, such as English, Mathematics, History and Economics. A wide range of Unit Standards were then allocated to each subject with some Unit Standards allocated to a number of subject areas. In the calculation of that subject, more weight was given to Excellence and Merit than Achieved in Achievement Standards, with an equal weight given to a pass in a Unit Standard and Achievement Standard. The final calculated score of sum (result x credit value) defined the discipline discrimination score for each subject domain for Years 11, 12 and 13. A Level 2 subject had slightly higher weight if done in Year 11, and so on. On average, most students attempt 18-20 credits per academic subject and, as determined by the real data, an average student would be expected to pass that number of credits in each subject. A very able student would be expected to get between a Merit and Excellence for that same number of credits.

Using this weighted discrimination score as a base for grouping standards within a subject discipline, the CEM Centre compared patterns between and within these subjects. For example, in 2005, 16,747 Year 9 students sat a baseline test of developed abilities under strict examination conditions. The results were used as an anchor to compare Year 11 NCEA subject areas (Figure 1).

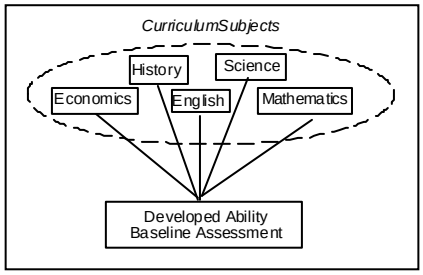

Figure 1: Diagram depicting how subjects were compared and monitored by the CEM Centre

The Year 9 common baseline has proven to be a good predictor of curriculum subjects and covers a range of skills that include vocabulary, basic mathematics skills, logic skills, visual processing skills, speed and accuracy. The same baseline is administered each year to large numbers of Year 9 students throughout New Zealand at the same time of the year, each assessment section is timed and consistent marking is done by the CEM Centre.

This common baseline defines the initial ability of the students, from which their Year 11 results can be compared. This approach gives a comparative view of a wide range of subjects.

Results - Within Subject Differences

As a result of the Year 11 subject analysis, each student has a Year 9 baseline score of developed abilities ("MidYIS Score") and a Year 11 discrimination score for that subject ("Year 11 Score"). Each dot on the scattergram in Figure 2 represents a student's Year 9 and Year 11 results. The line through the middle of the data is the national line of best fit (regression line) that shows the most likely Year 11 discrimination score for each initial Year 9 developed ability score. In a justification of the approach, Professor Peter Tymms from the University of Durham CEM Centre found that the errors between this ordinary least squares (OLS) approach and a multilevel modeling approach (MLM) were not different for most schools although there could be slightly higher error with OLS for small schools [referenced in (Fitz-Gibbon, 1997) Annex D].

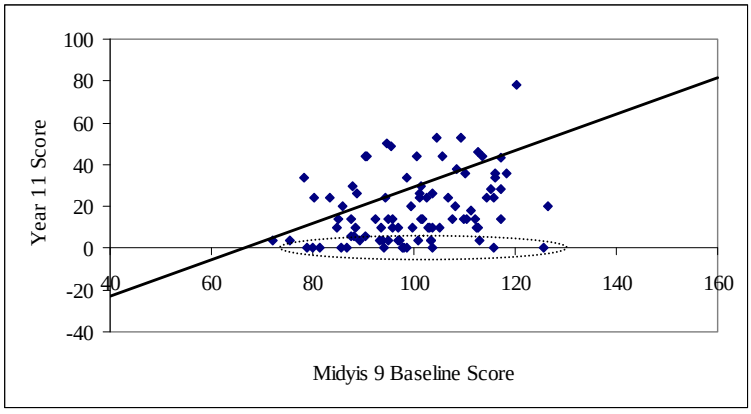

Figure 2: Scattergram showing the Year 11 scores for each Year 9 developed ability score in Science for one school

This approach to focusing on subjects highlights issues that may not be obvious under a generic system. By focusing on a generic 80 credits, it is possible for students to fail all standards in a particular subject area, that is, gain a Year 11 score of zero in that subject. Figure 2 illustrates this by showing the results from one school in Science. Students depicted as lining the X axis in Figure 2 are students who were registered for at least 14 credits in Year 11 Science, remained in the school but did not pass any Science standards. Although most subjects contained data points similar to these zero results, in the 2007 sample students who did both the CEM baseline in 2005 and Year 11 NCEA in 2007 it was especially noticeable with 5% of the Science students gaining a zero result in contrast to 0.3% in Mathematics and close to 0% in English. Unless the school is vigilant, results such as these can be overlooked and subsequently have negative implications for Year 12 teaching and learning if that student wishes to continue with standards in that subject area.

The lack of motivation in perceiving the 80 credits as a minimum can also be reflected in the data. Figure 3 illustrates that average and above average students with a baseline score from 90 to 120 (circled) register for 20 credits in a subject area but can gain fewer credits than a below average student. This is usually seen with an accumulation of more students below the regression line than expected.

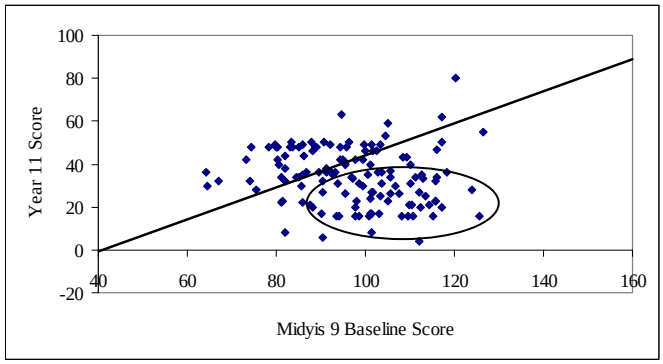

Figure 3: Scattergram showing the Year 11 scores for each Year 9 developed ability score in Mathematics for one school

Anecdotal evidence from schools points to a range of motivation issues that includes students not turning up to examinations or not answering questions once they were sitting the examination paper.

Results from the CEM Centre analysis are able to be used by schools to help motivate these students. The Achievement Goal Theory is used as a basis to focus students on a self-regulating mastery approach of internal target motivation rather than being motivated by competing with others (Dweck & Leggett, 1988; Nichols, 1984). Research shows that mastery goals can effectively forecast academic performance outcomes (Grant & Dweck, 2003; Young, 2007). Although it is too early to analyze the results, schools are starting to use the CEM Centre data to set subject targets for each student based on initial developed ability. This is done by finding the expected value from each baseline score using the latest regression lines. Based on expectation from initial developed ability, students are given a minimum target score for each subject. For example, a target score of 40 is expected of an average student in Mathematics (see Figure 3), representing a pass of 20 credits in that subject. This takes the focus away from a generic 80 credits and concentrates on achieving what is expected of a student in individual subject areas.

Results - Between Subject Differences

Measuring comparability of standards (or groups of standards) between subjects is statistically problematic because of the underlying assumptions, differences in subject-specific motivation and differences in subject complexity (Newton, 1997). The NCEA focus is on the number of standards passed and accumulation of standards

within a subject area are equally problematic for comparisons. However, using the regression lines for different subjects measured from the same baseline data, the means and regression lines can show consistent patterns over time that can be hidden within a generic standards-based system. Figure 4 is an example of regression lines for Year 11 (predominantly NCEA Level 1) subject area.

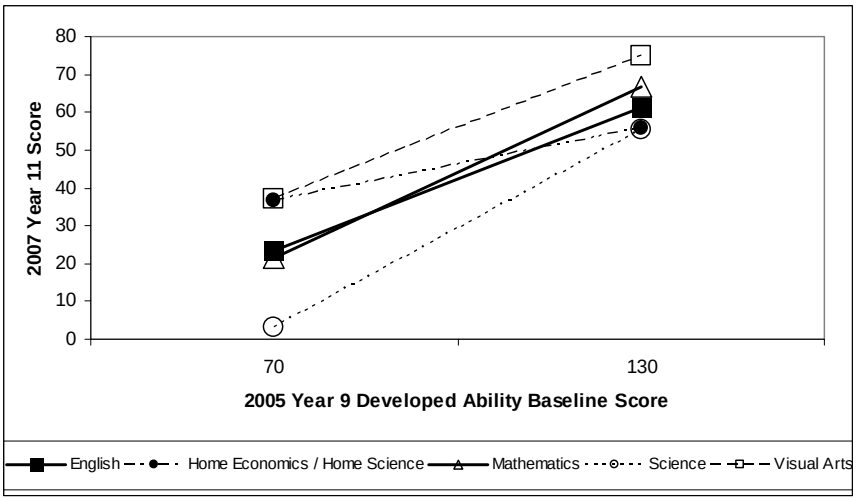

Figure 4: Regression lines for 2005 Year 9 to 2007 Year 11 progress in five subject areas.

There are definite differences between subjects, indicating that Meyer's 2007 survey where some students perceived a demotivation in some subjects areas, is supported by data. The estimated weighted Year 11score for Visual Arts is higher than all other subjects no matter what the ability level of the student (Figure 4). This indicates that students taking Visual Arts are likely to gain more credits, including Merit and Excellence results, than other subjects. Figure 4 also shows that the subject with the least discrimination between top and bottom scores is Home Economics with little more than a 5 credit difference between top and bottom students. English and Mathematics have similar discrimination with low ability students passing approximately 10 credits (a Year 11 score of '20') and high ability students expecting to score an approximate Merit average on 20 credits (Year 11 score of '60').

A large difference is in Science with the average weighted score being similar for the very able students (baseline score of 130), but considerably lower for the above average, average and below average students. Similar patterns emerged with the Year 13 analysis.

Using the multivariate analysis of variance (MANOVA), followed by individual analysis of variance (ANOVA), there was homogeneity of the variance-covariance matrices, but the means in Figure 4 were significantly different between subjects ($p<0.05$). Differences between subjects could have implications for student subject selection and potential career direction. There are also implications for the comparability of standards across subjects and standard options available for less able and average students.

Summary

In New Zealand secondary schools, a standards-based assessment system replaced the old norm-referenced system with a better a pass rate but introduced new concerns including student perception of demotivation in some subjects.

Success in obtaining NCEA qualifications under the more complex generic standards-based assessment system does not necessarily reflect success in specialized subjects. While specialization is not so important at the junior secondary level and below, specialization at the senior level is important for career preparation, the positive link between teacher specialization and achievement, scholarships and tertiary preparation.

Defining subjects within a generic standards-based system involves inclusion of appropriate standards and flexibility of overlapping standards. A common assessment of developed abilities, sat by a large number of students at the same time under the same strict examination conditions, allows the CEM Centre to compare within and across curriculum subjects without having to find ways to discriminate between standards that, by themselves, may not discriminate between abilities.

Using the value added OLS regression approach, the New Zealand CEM Centre found patterns within and between subjects. For example, students can gain NCEA by passing 80 credits but, at the same time, fail every standard they register for within a specialized subject area. Science stood out from the other subjects as having the highest number of zero results where students, registered for at least 14 credits of Science, did not pass even one Science standard. This has implications for

preparation into the next level if a student wishes to specialize in a science-related career or tertiary study.

Another pattern within subjects showed that too many average and above average students passed the same if not fewer standards than the below average students. By doing an achievable course of internal Unit Standards, lower ability students can pass a large number of credits which implies increased motivation. However, in too many schools, average and above average students gain fewer credits than the lower ability students. Often these higher ability students are doing a mixture of Achievement and Unit Standards. Some of the Achievement Standards are external examinations which are not always passed, especially if a student fails the questions that carry the 'achieved' weight, if the student does not turn up to the examination, or does not attempt the questions because (s)he has already passed the required 80 credits. The patterns point to demotivation for students who may have had a good chance of passing national examinations under the old norm-referenced system.

The 2005/2007 analysis also showed differences between subjects. At all ability levels, Visual Arts students tended to gain more credits, merits and excellences on average than in any other subject. There was also less discrimination between top and bottom students in Home Economics than in most other traditional subjects. Science had the largest discrimination between the top and bottom students. While top Science students achieved equally well as in most other subjects, bottom students struggled to pass Science standards and the average students did not pass as many standards with the same grades as in other subjects. If standards of equal credit weight are equivalent, the question is why the data shows that equivalent ability students can gain more credits in some subjects and fewer credits in other subjects.

Schools are taking note of the CEM Centre analysis. One implementation currently being attempted by a number of schools registered for CEM Centre projects is individualized subject targeting for each student where the focus is taken away from the generic 80 credits and back onto the subject. By using the student's original baseline result and working out an expected discrimination score for each broad subject area, a student is given an individualized target score for that subject. Throughout Year 11, the student uses this minimum target score to monitor their own

achievement as they approach their final NCEA qualifications. This subject focus not only aims to motivate the student to achieve success in specialized subject areas but it should automatically lead to an accumulation of the minimum 80 credits required for NCEA.

Comparative subject issues appear to have relevance to student motivation, one of the major causes of public concern with New Zealand's standards-based assessment system. The CEM Centre analysis is providing the platform for secondary schools to respond to this concern.

References

Black, M. S. (2000) *Old and Older: Curriculum Standards for History and Archeology*. Paper presented at the Annual Meeting of the American Education Research Association, New Orleans.

Dweck, C. S. & Leggett, E. L. (1988) A social-cognitive approach to motivation and personality, *Psychological Review, 95*(2), 256-273.

Fitz-Gibbon, C. (1997) *The Value Added National Project: Final Report*. (UK, CEM Centre, University of Durham).

Grant, H. & Dweck, C. S. (2003) Clarifying achievement goals and their impact, *Journal of Personality and Social Psychology, 85*(3), 541-553.

Heck, R. H. (2007) Examining the relationship between teacher quality as an organizational property of schools and student achievement growth rates., *Educational Administration Quality, 43*(4), 399-432.

Meyer, L., McClure, J., Walkey, F., McKenzie, L. & Weir, K. (2008) *Final Report: The Impact of the NCEA on Student Motivation* (Wellington, Victoria University of Wellington).

Ministry of Education. (2007) *The New Zealand Curriculum*. (Wellington, Ministry of Education).

Newton, P. E. (1997) Measuring comparability of standards between subjects: Why our statistical techniques do not make the grade, *British Educational Research Journal, 23*(4), 433-450.

Nichols, J. G. (1984) Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance, *Psychological Review, 91*(3), 328-246.

Young, J. W. (2007) Predicting college grades: the value of achievement goals in supplementing ability measures, *Assessment in Education: Principles, Policy & Practice, 14*(2), 233-249.