Standards-Based Assessment in the Senior Secondary School: A Review of the Literature


PETER RAWLINS, JILL BRANDON, JAN CHAPMAN, LINDA LEACH, GUYON NEUTZE, ADÈLE SCOTT, NICK ZEPKE
Massey University

BACKGROUND

In 2004, the New Zealand Qualifications Authority (NZQA) completed its implementation of a standards-based assessment (SBA) system with the National Certificate of Educational Achievement (NCEA) as the main qualification for secondary students. NZQA is conducting ongoing research into SBA practices in the senior secondary school. To ensure that findings from this research build on and link to international and national literature, a team from Massey University were commissioned to conduct a review of the existing literature.

Discussion with NZQA staff resulted in the identification of five key themes: policy issues, technical matters, teaching; learning, and diversity. A search strategy using international databases and a set of inclusion/exclusion criteria was developed. The review team sought to identify articles about SBA and relevant to the senior secondary school. It was decided to include not only empirical studies but also literature reviews and theoretical work, and to focus on more recent work, particularly that done since the mid 1990s.

The initial search located 90 pages of bibliographic information about studies of interest. From this list those studies that seemed most closely related to the themes were selected and accessed then, the inclusion/exclusion criteria were applied to them. The search iteration continued for over four months until the writing of the critical review of the studies began.

From this 130 items initially deemed suitable, 88 publications were selected for analysis and 80 were selected for the synthesis of the literature. An annotated, evaluative, bibliography was completed by developing a template format that summarises key information from each article reviewed. These templates were then used to write a critical literature review. This paper summarises some of the key findings of the full literature review.

One significant problem facing the review team was that SBA is difficult to define (Croft, 1993). The term has been used in numerous ways to mean different things, even within the NZQA literature. As an example, criterion-referenced assessment has been used synonymously with criteria-based assessment, standards-based assessment, standards-referenced assessment, competency-based assessment and achievement-based assessment. On other occasions these terms have had different meanings. Competency-based assessment and achievement-based assessment are frequently interpreted as different versions of SBA. In this review, all paradigms have been considered in the synthesis of the literature.
INTRODUCTION

The debates surrounding the value of SBA are complex. There are avid proponents of this type of assessment and equally passionate critics. Advocates stress improved transparency and understanding of the assessment process (Barker in Peddie & Tuck, 1995; Francisco, 1999; Tomlinson, 2002); higher levels of student achievement (Supovitz, 2001); improved links between knowledge and performance (Barker in Peddie & Tuck, 1995); improved generic skills (Gfroerer, 2000); more stability and robustness of teacher judgements from diverse assessment methods (Pitman, 1985); enhanced international comparability (Peddie & Tuck, 1995); and the potential democratisation of learning and the erosion of traditional barriers and quotas (Barker in Peddie & Tuck, 1995). Indeed, Gipps (1994) argues that SBA ameliorates competition, reduces anxiety, increases intrinsic motivation, and promotes achievement, cooperation, self-efficacy, metacognition and deep learning.

Critics are equally vocal in their opposition to SBA. Lee and Lee (2000) and Sizmur and Sainsbury (1997) identify issues of proliferation, atomisation and specificity in SBA as resulting in manageability and workload problems for teachers. Dearing (cited in Sizmur and Sainsbury, 1997) states that ‘many teachers feel that the mechanics of recording teacher assessment information have interfered with teaching and learning’ (p.137). Singh-Morris (1997) contends that standards are reductionist. Students do not learn in discretely defined bits so assessing pre-specified skills may lead to a narrowing of the curriculum, over-assessment and the growth of an assessment curriculum rather than a learning curriculum.

SELECTED POLICY ISSUES IN STANDARDS-BASED ASSESSMENT

The literature indicates that SBA is a dominant assessment paradigm in English speaking countries (Gipps, 1994; Philips, 1998; Strachan, 2002; Tognolini et al., 2001; Wolf, 1995). Indeed SBA has been theorised as just one aspect of a wider educational reform movement (Bennett & Merrick, 2004; Cowan et al., 2002; WestEd, 2000). Clune (2001) takes up this view by proposing a holistic model of standards-based reform, which through purposeful activities leads to standards-based policy, which then leads to a rigorously implemented standards-based curriculum for all students. This, in turn leads to high student achievement through standards-based assessment.

A distinctive feature of the New Zealand migration of assessment policy was the decision to create a single, National Qualifications Framework (NZQF). Philips (2003) considers the NZQF to be possibly the most comprehensive in the world, embracing both academic and vocational qualifications. Initially the NZQF, as framed by NZQA, consisted of a potentially infinite number of vocational and academic Unit Standards, sorted into content domains and eight levels of difficulty that would span from the senior secondary level through to the tertiary level. However, this unitary framework did not find universal favour, with the following three aspects attracting criticism. Firstly, Unit Standards are more suited to assessment of technical or practical skills than higher order thinking skills. Secondly, they are unsuited to university learning. Finally, their use has the potential to atomise or fragment integrated knowledge and thereby distort
the curriculum (Codd et al., in Peddie & Tuck, 1995; Elley & Hall in Peddie & Tuck, 1995; Hearn, 1997; Lee & Lee, 2000).

This ongoing critique resulted in the ‘broadening’ of the NQF in 1996 (Lee & Lee, 2000), and in the announcement of the NCEA in the Achievement 2001 Policy. This saw the introduction of a new standards based assessment tool, Achievement Standards, that allowed for graded achievement levels – not achieved; achieved; achieved with merit, and achieved with excellence – for both the internally and externally assessed Achievement Standards (Lee & Lee, 2000; Philips, 2003).

Standards-based assessment, it has been claimed, is costly to implement (Gilmore, 1991; Linn & Herman, 1997; Wolf, 1995). It is time-consuming in absolute terms, for both candidates and assessor and it imposes major costs in terms of equipment, provision of wide-ranging assessment situations, and repeated evidence of mastery. Accordingly, the provision of adequate resourcing must be a major policy decision for any organisation seeking to successful implement SBA.

TECHNICAL CONSIDERATIONS IN STANDARDS-BASED ASSESSMENT

There is considerable debate in the literature about which of the purposes of assessment should take priority – formative, summative or evaluative assessment. Some argue that public confidence in the system of assessment is vital and therefore the evaluation and accountability aspect is the most important (e.g., Hall, 1999; Winch & Gingell, 1996). Others critique the prevalence of assessment for accountability as it leads to ‘teaching to’ and ‘learning for’ the test (Boss et al., 2001; Willan, 2004).

A number of authors, particularly in the New Zealand context, critique the atomisation of learning in SBA (Barrington, 2004; Boss et al., 2001; Davis, 1996; Hall, 1999; Hearn, 1997; New Zealand Post Primary Teachers Association (NZPPTA), 2002; Peddie & Tuck, 1995; Strathdee & Hughes, 2001). Holistic knowledge and understanding gives way to knowledge that is more easily measured at the expense of critical, creative and integrated thinking. Davis (1995) argues that rich knowledge and skills are relational to many other things involving complex and holistic systems of belief connected in diverse ways. Learning does not occur in discrete bits but is an integrated process. Accordingly, SBA may run the risk of fostering a ‘bricks without mortar’ approach to course design, delivery and assessment (Hall, 1999). In opposition to this view, Hearn (1997), reporting NZPPTA views, argues that atomisation is not an inevitable consequence of SBA and that it can elicit sophisticated skills and knowledge and does not inhibit them. Hager et al. (1994) recommend an integrated approach to programme design in which discrete standards are combined to better represent the field of knowledge.

There has also been much debate in the literature regarding validity and reliability in SBA. Validity and reliability presume to guarantee, on the one hand, that assessments are fit for the purposes they are set (Gipps, 1994; Hall, 1999), and on the other, that these purposes are consistently met. Crombie (in Peddie & Tuck, 1995) warns that concern with reliability can lead to an over-restrictive view of what is being tested, and may result in invalid assessment (Linn & Herman, 1997). Using the notions of consequential validity (Gipps, 1994) and ecological validity (Black & Wiliam, 1998) the balance of opinion in the literature seems to be that SBA can be valid as long as it assesses course learning outcomes and is fit for the context for which it is intended. Gipps
(1994) also contends that SBA cannot be responsive to traditional reliability
criteria. She argues instead that, since a set of standards should represent
consensus among stakeholders on what students must know and be able to do,
moderation by consensus groups and external consultations, rather than
statistical moderation, is more appropriate. In support of this view, Hager et al.
(1994) point out that informed professional judgement has been found typically
to have a high level of reliability.

Alignment is a key technical matter addressed in the literature. The
concept has different constructions. It may specify alignment between
standards and assessments, standards and teaching, standards and
curriculum, and between teaching and assessments. Any or all of these may be
present in specific cases. The evidence is that the closer the alignment between
these factors, the better the students achieve (e.g., Clune, 2001; Linn &
Herman, 1997; Porter & Smithson, 2001).

THE IMPACT OF STANDARDS-BASED ASSESSMENT ON TEACHING

Standards-based reform is consistent with an emerging view of
assessment for learning rather than assessment of learning (Assessment
Reform Group, 1999; Black & William, 1998; Crooks, 1998). This paradigm calls
for deep changes both in teachers’ perceptions of their own role in relation to
their students and in their classroom practice. In particular, it suggests a move
to a more student-centred pedagogical approach, placing the student in a more
active role in the learning, teaching and assessment cycle, thus creating a
partnership between student and teacher. The clarity and transparency of
assessment standards help teachers provide students with information of what
they know and can do and, more importantly, a clear picture of what they need
to do to improve so they can take charge of their own learning (Black & William,
1998; Crooks, 1988).

The balance of the literature suggests that the impact of SBA on pedagogy
is moderately positive. Studies reporting teachers’ concerns that assessment
was impinging on teaching time and becoming more important than the ‘joy of
learning’, (Boss et al., 2001; Preece & Skinner, 1999) are balanced by studies
that found the effects were not extreme and in essence teaching was not being
‘re-invented’ in the image of assessment (Wilson & Floden, 2001). Other
reports found that SBA had a positive impact on pedagogy (Barrington, 2004;
Bushnell, 1992; Kannapel et al., 2001) with the clarity of assessment standards
creating a catalyst for teachers using their professional judgement to create a
more coherent teaching practice embracing the old and the new (Wilson &
Floden, 2001), resulting in pedagogy that is more organised, systematic and
standardised (Preece & Skinner, 1999) and creating more student-centred
approaches (Eng, 1992).

As mentioned earlier, one of the criticisms of SBA is that it atomises
the curriculum and fragments intricately integrated knowledge (e.g., Peddie & Tuck,
1995). The literature is divided on this issue. Some studies concluded that
assessment per se has a constricting effect on curriculum and pedagogy with
teachers tending to teach for assessment rather than for learning (Harlen &
Crick, 2003; Preece & Skinner, 1999). Other studies have reported that SBA
has the effect of expanding curriculum in some schools into areas that had
previously received little attention, for example, arts and humanities (Kannapel
et al., 2001). A study by Wilson and Floden (2001) found that, although every
teacher reported that tests affected instruction, independent classroom observations showed that assessments were, in fact, neither predominant nor entirely absent, suggesting that teachers' perceptions are not always accurate.

Studies by Bushnell (1992) and Francisco (1999) confirm that, although the process of implementing SBA was initially time consuming, the benefits to improved assessment and the re-evaluation of teaching strategies, outweighed the cost. Improvement to professional dialogue between teachers has been reported as the most productive and meaningful professional development in recent times (Clune, 2001; Harlen & Crick, 2003; James, 2000; Wilson & Floden, 2001).

THE IMPACT OF STANDARDS-BASED ASSESSMENT ON LEARNING

The balance of evidence from the literature suggests that standards-based reform, and its implied pedagogical changes, have a positive impact on student learning and achievement (e.g., Black & Wiliam, 1998; Bushnell, 1992; Clune, 2001; Gipps, 1994; Hipkins, 2004; Hipkins et al., 2004; Kannapel et al., 2001; Ronis, 1999). There are, however, some studies that report that the effects may only be modest (Supovitz, 2001) or not unequivocal (Khattri et al., 1998). In addition, there are some indications that SBA may create a performance orientation in students which is contra to the very philosophical underpinnings of standards-based reform (e.g., James, 2000; Stefanou & Parkes, 2003).

Given that assessment in New Zealand is no longer about students competing against one another, but about achieving to a set standard, it is not surprising that research suggests that students are becoming more empowered by SBA and are increasing in confidence (Barrington, 2004; Hipkins & Vaughan, 2004). Many schools in New Zealand are beginning to acknowledge the value of SBA to learning (Barrington, 2004; Gibson, 2004; Mallard, 2004).

The potential for quality formative assessment is a key feature of SBA that can lead to significant improvement in student achievement (e.g., Assessment Reform Group, 1999; Black & Wiliam, 1998; Francisco, 1999; James, 2000; Supovitz, 2001). By providing clear objectives, students can recognise gaps in their learning between the current level and desired level and can take effective action to close the gap (Brookhart, 2001; Ronis, 1999). In addition to this, SBA contributes to an increased understanding of the assessment process and a shift from quality control in learning to quality assurance (Black & Wiliam, 1998).

IMPACT OF STANDARDS-BASED ASSESSMENT ON DIVERSITY

The academic achievements of diverse learners within SBA systems have been mixed. The research suggests that although diverse students perform better under SBA than under a norm-referenced system, there is still a significant gap between the achievement of students with special needs and their middle class majority peers (Ortiz, 2000), particularly minority students and those from low income households (Kannapel et al., 2001; Madaus & Clarke, 2001). It remains to be seen whether NCEA, as a high-stakes national assessment system, can provide more positive outcomes for diverse students than the previous system, although Mallard (2004) has indicated that there are initial signs of improved achievement by Māori and Pasifika students.

Standards-based assessment potentially provides schools with greater opportunity to adapt assessment tasks to meet the needs of diverse learners, while still assessing the set standard(s) (Hager et al., 1994; Hipkins et al.,
2004). In order that SBA is fully inclusive there are a number of issues that need to be addressed. In particular, the provision of accommodations such as more time and special aids (Thurlow, 2000), and alternative assessments, designed and moderated, so they are viewed as comparable and of equal status to those commonly used.

CONCLUSION

The standards-based reform movement and SBA have been the subject of much debate in the international literature. Although it should be recognised that there are ongoing issues of manageability, validity and reliability, alignment to, and atomisation of, the curriculum, it should also be recognised that SBA has a positive impact on teacher pedagogy, student achievement and programme adaptation to better meet the needs of an increasingly diverse range of students.

REFERENCES


The opinions expressed are those of the paper author(s) and not the New Zealand Journal of Teachers’ Work. Copyright is held by individual authors but offprints in the published format only may be distributed freely by individuals provided that the source is fully acknowledged. [ISSN-1176-6662]
About the Author(s)

New Zealand Journal of Teachers’ Work, Volume 2, Issue 2, 2005

PETER RAWLINS, JILL BRANDON, JAN CHAPMAN, LINDA LEACH, GUYON NEUTZE, ADÈLE SCOTT, NICK ZEPKE

Massey University

Nick Zepke is an Associate Professor in the School of Educational Studies and coordinator of the undergraduate programmes in adult education. Dr Linda Leach is a senior lecturer in the School of Educational Studies and the coordinator of postgraduate programmes in adult education. Jill Brandon is the Head of the School of Arts, Development and Health. Jan Chapman is a senior lecturer with the School of Curriculum and Pedagogy. Guyon Neutze was formerly a senior lecturer in the Department of Social and Policy Studies in Education. Peter Rawlins is a lecturer in the School of Curriculum and Pedagogy. Adéle Scott is a senior lecturer in the School of Educational Studies.

Peter Rawlins
Massey University College of Education
Private Bag 11 222
Palmerston North
p.rawlins@massey.ac.nz