



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

*School of Education*

Te Kura Toi Tangata

**Waikato Journal of Education**

Te Hautaka Mātauranga o Waikato

Volume 14: 2008/2009



# WAIKATO JOURNAL OF EDUCATION TE HAUTAKA MĀTAURANGA O WAIKATO

**Editor:**

Rosemary De Luca

**Special Section Editor**

Clive McGee

**Editorial Board:**

Karen Barbour

Beverley Bell

Toni Bruce

Margaret Carr

Bronwen Cowie

Deborah Fraser

Richard Hill

Rangimarie Mahuika

Judy Moreland

Sally Peters

Clive Pope

Lesley Rameka

Noeline Wright

*Waikato Journal of Education* is a refereed journal, published annually, based in the School of Education, University of Waikato, Hamilton, 3240, New Zealand. It publishes articles in the broad field of education. For further information visit the WJE website <http://www.soe.waikato.ac.nz/wje/>

Correspondence and Books for review should be addressed to: Research Manager, Wilf Malcolm Institute of Educational Research, School of Education, Private Bag 3105, The University of Waikato, Hamilton, 3240, New Zealand. Email: [wmier@waikato.ac.nz](mailto:wmier@waikato.ac.nz)

Business correspondence: Orders, subscription payments and other enquiries should be sent to the Administrator, *Waikato Journal of Education*, Wilf Malcolm Institute of Educational Research, School of Education, The University of Waikato, Private Bag 3105, Hamilton, 3240, New Zealand, Email: [wmier@waikato.ac.nz](mailto:wmier@waikato.ac.nz)

Subscriptions: Within NZ \$35; Overseas NZ \$45

Copyright: © School of Education, The University of Waikato

Publisher: School of Education, The University of Waikato

Cover design: Donn Ratana

Printed by: Waikato Print

ISSN 1173-6135

# WAIKATO JOURNAL OF EDUCATION

TE HAUTAKA MĀTAURANGA O WAIKATO

VOLUME 14, 2008/2009

<b>Editorial</b> ROSEMARY DE LUCA	3
<b>Not Empty Vessels: New Zealand Pre-Service Additional Language Teacher Identity</b> NICOLA DALY	5
<b>Searching for Standards in the NCEA: Assessing Musical Performance</b> GRAHAM MCPHAIL	15
<b>One Hundred Years of Sylvia Ashton-Warner: An Introduction</b> SUE MIDDLETON	31
<b>What's in a Word?</b> MARILYN BARLOW	35
<b>Finding Dance in Sylvia's Classroom</b> ADRIENNE SANSOM	47
<b>The Musician in the Classroom: Sylvia and a Pedagogy of Artistic Knowing and Meaning-Making</b> TREVOR THWAITES	57
<b>The Development, Validation and Application of a Science Curriculum Delivery Evaluation Questionnaire for Indigenous Māori Settings</b> BRIAN LEWTHWAITE AND ANARU WOOD	69
<hr/>	
<b>Introduction to the Special Section on Curriculum</b> CLIVE MCGEE	89
<b>The Context of Contemporary Curriculum Change</b> CLIVE MCGEE AND BRONWEN COWIE	91
<b>Passionate and Proactive: The Role of the Secondary Principal in Leading Curriculum Change</b> MICHELE MORRISON AND BEVERLEY COOPER	105

<b>The Role of Professional Development and Learning in the Early Adoption of the New Zealand Curriculum by Schools</b> JENNY FERRIER-KERR, PAUL KEOWN AND ANNE HUME	<b>123</b>
---	------------

---

<b>The Risk and Resilience Framework and its Implications for Teachers and Schools</b> PETER STANLEY	<b>139</b>
---	------------

# SEARCHING FOR STANDARDS IN THE NCEA: ASSESSING MUSICAL PERFORMANCE

**GRAHAM MCPHAIL**

*Doctoral student, Faculty of Education, University of Auckland.*

**ABSTRACT:** *In this paper it is argued that the theory and practice of standards-based assessment within the context of the National Certificate of Educational Achievement (NCEA) has not been clearly articulated for teachers. The difficulty of specifying and promulgating standards in appropriate forms and the lack of clarity present in the support materials and training provided for teachers are examined. Through the analysis of an internally assessed Music Achievement Standard currently available in the NCEA, it will be argued that standards can be neither definitively described nor easily assessed, but that a credible standard is reliant on a number of components. It is the combination of these components that is significant if standards are to function effectively in summative contexts, particularly for high stakes national qualifications. The support materials and training music teachers received during the introduction of the NCEA lacked clarity and this has resulted in a weak link in the chain of components required for a robust assessment system. Teachers need access to quality support materials and the opportunity for on-going professional development in relation to standards-based assessment.*

**KEYWORDS:**

NCEA, music assessment, assessment, standards, standards-based assessment

**INTRODUCTION**

Standards-based assessment in New Zealand secondary school education has emerged in a context of some controversy, particularly in relation to the National Certificate of Educational Achievement (NCEA). Its proponents suggest that this form of criterion-referenced assessment should be supported because of its “lucid and unambiguous descriptions of what is to be measured, with the learner’s performance being interpreted by reference to specific standards, which are set prior to instruction” (Peddie & Tuck, 1995, p. 9). The ideal of measuring students against standards rather than each other marks a radical departure from the norm-referenced systems that have dominated assessment in New Zealand schools. For this reason alone many would argue that the system is an improvement; but others have objected to standards in education for political, philosophical and pedagogical reasons. Its detractors see the emergence of the objectives movement as part of a positivist, reductionist, and market approach that is driven primarily by values of accountability and managerialism rather than educational concerns.<sup>1</sup> National standards can appear to present a reductionist view of education, particularly where

they “*become* the curriculum” rather than “promoting accountability or ‘quality control’ *for* curriculum” (Regelski, 2004, p. 256).

Despite such reservations, standards do have the potential to provide positive educational experiences and outcomes when utilised as tools for teaching and learning (Elley, 1995). The possibility for standards to provide a reliable and unambiguous measurement, particularly in high stakes qualifications, is more problematic. Through the analysis of an internally assessed Music Achievement Standard (Level 2 solo performance) currently available as part of the NCEA, it will be argued that standards can be neither unambiguously described nor easily assessed, but that a credible standard is reliant on a number of components. It is the combination of these components that is significant if standards are to function effectively in summative contexts, particularly for high stakes national qualifications. These components include the published Standard, with its criteria and explanatory notes, exemplars that act as benchmarks for the standard, a moderation process that supports inter-school comparability, training for teachers in assessment and task design, and teachers’ professional knowledge.

The first part of this paper considers the difficulty of specifying and promulgating standards in appropriate forms and assessing against them, arguing that verbal description alone cannot specify a standard. Reference to external factors to anchor the standard is required. Sadler (1987) has noted that verbal descriptors and exemplars taken together with teacher training provide the most promising framework for standards-based assessment. Given the importance of each component in this view, the second part of the paper argues that the guidance provided during NCEA training days lacked clarity, particularly in relation to the development of teachers’ abilities in making qualitative judgments. This is exemplified by the ambiguous status afforded the music exemplar materials within the assessment process.

### **SEARCHING FOR MUSICAL STANDARDS IN THE NCEA**

The development of individual performance skills occupies a significant place in the Western classical music traditions that have dominated New Zealand music education. This is reflected in the weighting given to solo performance Achievement Standards in the NCEA.<sup>ii</sup> The title of the Level 2 solo performance Standard 2.1 (AS90264, version 2) (New Zealand Qualifications Authority, 2004) is self-explanatory: “Present contrasting performances as a featured soloist”.<sup>iii</sup> This title defines the task involved, but it gives no indication of a standard of attainment other than that inherent in the NCEA Level number.

The achievement criteria for this Standard outline six broad distinguishing properties or characteristics of the performance that will guide the assessor and indicate whether the student has succeeded in meeting the level of achievement required for the Standard. These criteria are contrast, accuracy, technical skills, musicianship, presentation skills and fluency (see Figure 1).

**Figure 1. New Zealand Qualifications Authority (2004). Achievement Standard 90264 (version 2).**

**Achievement Standard**

<b>Subject Reference</b>	Music 2.1		
<b>Title</b>	Present contrasting performances as a featured soloist		
<b>Level</b>	2	<b>Credits</b>	5
		<b>Assessment</b>	Internal
<b>Subfield</b>	Music		
<b>Domain</b>	Making Music		
<b>Registration date</b>	20 October 2004	<b>Date version published</b>	20 October 2004

---

This achievement standard involves the presentation of contrasting performances of music by a featured soloist.

**Achievement Criteria**

Achievement	Achievement with Merit	Achievement with Excellence
<ul style="list-style-type: none"> <li>Present generally accurate contrasting performances, demonstrating a range of technical skills, and appropriate musicianship and presentation skills.</li> </ul>	<ul style="list-style-type: none"> <li>Present fluent and mostly accurate contrasting performances, demonstrating a wide range of technical skills, and effective musicianship and presentation skills.</li> </ul>	<ul style="list-style-type: none"> <li>Confidently present fluent and highly accurate contrasting performances, demonstrating secure techniques, and convincing musicianship and presentation skills.</li> </ul>

(New Zealand Qualifications Authority, 2004)

The criteria of contrast, accuracy, fluency, and technical skills are relatively ‘sharp’ in definition when compared to the less concrete aspect of musicianship (see below), but nevertheless they present problems of interpretation. For example, accuracy is a dimension of a musical performance that can, in some cases, be judged by following a music score, but depending on the genre this may not always be applicable (as is the case in jazz, Baroque slow movements, or contemporary popular music, which can be largely or totally unscripted). If an explicit score does exist, accuracy could be gauged purely in terms of the number of correct notes, but this kind of atomistic approach does not reflect the way musicians assess a musical performance in normal circumstances. Accuracy could also refer to the realisation of the composer’s dynamics, articulations, and phrasing as marked on the score, in which case a much broader view of the criterion is being assessed. A potential contradiction could occur when certain details of the score might not be ‘accurately’ realised, but the result is nevertheless convincing and musical. Musicians work with this tension between the composer’s markings and appropriate creativity in every performance, and it may be that the authors of the Standard would expect teachers to bring this professional knowledge into play when making judgements in this area. What Sadler (1989) describes as latent criteria then come to the fore in the

assessment process. Assessors might trade-off aspects of the performance that are not in the explicit or manifest criteria but that nevertheless contribute to the work meeting the standard. For example, assessors may choose to acknowledge the greater importance of 'feel' over tone in a particular performance. In this regard, there is the potential for teachers to exhibit critical judgement in a form of extended professionalism rather than a technocratic reading of the standard.

The musicianship and presentation skills criteria are less concrete. To make judgements in relation to musicianship in particular is far more subjective, and calls into play the experience and connoisseurship of teachers; their tacit knowledge of what musicianship involves. This kind of tacit knowledge often resides "essentially in unarticulated form, inside the heads of assessors ..." (Sadler, 1987, p. 199) and "... provides the backdrop against which explicit knowledge can be interpreted and understood" (O'Donovan, Price, & Rust, 2004, p. 333). However, the aim of the verbal description of the Standard is to go some way towards making such aspects publicly accessible. The explanatory notes accompanying the Standard give some guidance: "Musicianship refers to musical awareness and understanding of the musical style. This includes phrasing, dynamics, rhythm and feel" (see appendix 2, exploratory note 6) (New Zealand Qualifications Authority, 2004). The interpretation of the explanatory notes demands a high level of teacher experience of musicianship.<sup>iv</sup>

In Achievement Standards, within each level of the achievement criteria, modifiers or verbal quantifiers are used to describe how the criteria can be performed with varying degrees of success or "the amount of each variable that is needed to decide whether or not a criterion has been met" (Raths, 1999, p. 137<sup>v</sup>); in this case to enable a judgement for achievement, merit or excellence to be made. The performance will be *generally* accurate, *mostly* accurate, or *highly* accurate; the performance will demonstrate *a range* of technical skills, *a wide range* of technical skills, or *secure* technical techniques and *appropriate*, *effective*, or *convincing* musicianship and presentation skills. *Confidence* also appears as an indicator of excellence (see Figure 1). At first glance such verbal descriptions appear quite straightforward, but in practice they are too vague or fuzzy to provide an objective reference for what the standard might actually be. Such descriptors actually introduce an element of normative comparison (particularly within the context of NCEA, which asks the assessor to define three standards within one: achievement, merit or excellence) but are certainly not in themselves an indication of an absolute standard. Qualifiers, such as "generally", "mostly", "highly", "confidently" and so on only take on meaning in relation to a context. Sadler (1987) notes: "The context obviously determines the relation between the quantity present and the use of the grading labels...a fuzzy standard presupposes a set of existential referents and cannot be understood without them" (p. 204). Furthermore, the same criteria, in basically the same format, appear at each of the three levels of the NCEA (see appendix 1) and therefore do not in themselves provide a concrete reference point or define a level of achievement. Sadler (1987) also notes that "it is often impossible to tell from a verbal description alone whether the standard is intended to refer to, say, middle secondary school or undergraduate level" (p. 204). The



criteria are fuzzy, being relative rather than absolute, and require a context to carry any actual meaning.

It is the explanatory notes that provide the explicit context for the Music Standard under discussion (see appendix 2). Buried in the seventh of eight explanatory notes is a significant statement that links the criteria to a stable reference point that teachers will find meaningful: “The performance should reflect the equivalent technical and musical demands of a fourth year of study through itinerant lessons” (see appendix 2, explanatory note 7) (New Zealand Qualifications Authority, 2004).<sup>vi</sup> This statement gives external formulation to teachers’ tacit knowledge or knowledge from the field (Sadler, 1987, p. 199). Experienced practitioners should know what can be reasonably expected of a student in his or her third year of musical instrument tuition through the public system, and be able to relate the achievement criteria accordingly. Thus the Standard cannot function effectively without this reference to “what is known to have happened so far (and therefore what is known to be feasible)...so as to correspond with what is achievable in the field” (Sadler, 1987, p. 197). The explanatory note provides the necessary fixed point of reference from which a level might be recognised and a judgement made.

The explanatory notes of the Standard provide important information and guidance. Whether they should be considered binding in relation to the fulfilment of the Standard is unclear from the written document, and this is a practical weakness in the format of the document for guiding practitioners. As mentioned earlier, the notes further define the criteria and help clarify the verbal descriptors. That an apparently clearly defined Standard should need a further page or more of an explanation adds weight to the notion that defining achievement-based criteria is not a simple procedure. O’Donovan, Price and Rust (2004) note that “...the construction of ever more comprehensive and precise anchor definitions quickly became self-defeating” (p. 327).

Apart from the use of qualifiers, no definite guidance is provided on how assessors might determine the threshold between merit and excellence, and at which point such levels might overlap with the adjacent levels. The criteria might be better considered as an attempt to represent three standards within one. Thwaites (2002) provides some guidance as to what the authors may have had in mind<sup>vii</sup>:

Sufficiency in achievement standards involves three levels of critical decision-making by the teacher and the question is: Is sufficiency in Excellence merely more of the same that was required for Achievement, or should Excellence require knowledge and skills with more breadth and depth? (p. 29)

It would appear then that assessors should be looking for extra qualities (highly, confidently, convincing) that might not be the ‘norm’ for a student in his or her third year of tuition. The process then seems to require aspects of both absolute and relative judgements. Yet again, such decisions rely on teachers’ knowledge from experience in the field, with the achievement criteria and accompanying explanatory notes providing a focus for the assessment process. There is no mention of the relative importance of the criteria or whether all aspects must be present in

equal weight (as is the case with Unit Standards, where ‘elements’ and ‘performance criteria’ require the teacher to make separate decisions). Thus the assessors are able to respond to the holistic impact of the performance. The criteria appear sensible and are representative of the way in which Western musical performance is generally approached and perceived in the wider musical world. In other words, the Achievement Standard generally reflects the practices of the ‘real world’, and relies on competent assessors to make professional judgements including trade-offs between manifest and latent criteria (Sadler, 1989).

Musical performance has traditionally been assessed in a configurational manner where assessors, at least initially, react to the work as a whole before justifying responses through specific criteria (Kaplan, as cited in Sadler, 1989). Mills (1991) discusses two approaches to assessing music performance, holistic and segmented, and refers to a study that demonstrates holistic assessment resulting in a greater consensus of results than assessment that was defined by strictly applied segmented criteria. Stanley, Brooker and Gilbert (2001) also found that fewer criteria expressed as statements of performance characteristics were preferred by assessors. They suggest a balance between criteria and holistic aspects of assessment, noting that criteria can aid in focusing assessment responses and thus potentially aid in reliability between examiners. The overly analytic approach cited by Mills (1991), where assessment was based on five practical and five interpretative criteria, has been avoided in this local Standard, which overall aims at a more holistic emphasis. Sadler (2007) warns that overly atomized approaches to assessment can create a focus on “discrete competencies, rather than on competence” (p. 392). The achievement criteria act as recurring variables that assist in making decisions of classification (Sadler, 1987; Raths, 1999) and essentially constitute what Grudnoff and Tuck (2003) describe as content standards: broad descriptions of skills that will be made explicit and concrete in the actual performance. The revised Level 1 Standard, AS90012, version 3 has an additional explanatory note: “The performance will be assessed holistically. This means considering the whole performance, rather than isolating small technical inaccuracies or minor slip-ups and placing emphasis on these rather than on the musicality of the performance” (Explanatory note 12) (New Zealand Qualifications Authority, 2008).

In a comparison of Unit Standards with Achievement Standards, Thwaites (2001) points out that the Achievement Standards “... set out to be more holistic, taking an overall view of the student’s performance” rather than giving detailed lists of what must be achieved as is the case with Unit Standards (p. 31). There is the danger that criteria expressed as distinct outcomes become “operationally isolated from the others ... and what students are supposed to acquire by way of integrated knowledge and skills recedes into the background” (Sadler, 2005, p. 181). Moreover, the Unit Standards for music performance suffer from the same limitations as Achievement Standards. Despite greater levels of criteria specification, they do not clarify the existence of a standard that can be adequately defined by verbal descriptors any more readily than Achievement Standards. The special notes in the Unit Standards also make reference to the years of study the student is likely to have undertaken, thus anchoring the Standard to an external

reference point: "... candidates for this unit standard will be completing three years of musical performance at secondary school ... or have reached an equivalent standard" (Special note, 1, US 10651) (New Zealand Qualifications Authority, 2008).

Criteria inherent in standards can be made concrete by the use of exemplars. Sadler describes exemplars as key examples chosen to illustrate typical levels of competence. They are not the standards themselves, but are indicative of them (Sadler, 1987). Such concrete examples should act as benchmarks and help teachers establish the standard. Over time teachers will develop a sense of where standards lie, but even experienced teachers may need to refer to benchmarks periodically to realign their judgments in difficult cases or in instances where the local quality of performance may vary significantly from that of the national standard. Sadler (1987) also suggests that over time exemplars will need revision since they "incorporate elements of fashion, cultural tradition, or current technology, virtually ensuring that they go gradually out of date" (p. 201). Sadler (1987) has noted that verbal descriptors and exemplars taken together with teacher training provide the most promising framework for standards-based assessment.

### **SUPPORTING THE STANDARDS – TRAINING AND EXEMPLARS**

The first part of this paper has argued that there are inherent issues surrounding standards-based assessment, particularly problems with verbal description of standards, external anchorage, and the need for exemplars. The second part of this paper will argue that the key to addressing these issues lies in training and in support materials. Teachers need to develop the relevant conceptual tools as well as have a clear overview of how standards, tasks, exemplars, assessment and moderation fit together to create a sound and defensible national assessment system. It is debatable whether these aspects have been clearly articulated either during the NCEA training days or subsequently.

Under the School Certificate system Music teachers were obliged to meet in school-based cluster groups twice in a year. Each teacher brought video evidence of a range of assessed performances and under the guidance of a local moderator results were adjusted as appropriate in relation to a previously set benchmark. Such meetings enabled professional development and interchange that is now missing from the NCEA system. Teachers now assess mostly in isolation (many Music teachers are sole charge), submitting a sample of internally assessed work for moderation<sup>viii</sup> rather than working in cluster groups to develop shared understandings in a community of practice for standards development. Teachers need to develop both an insider's and an outsider's role in the moderation process, recognising the likely tension between needing to have their professional views valued while accepting that the intention of moderation is to relate individual work to generalised standards (Radnor & Shaw, 1995).

The provision of exemplars, which consist of videoed student performances and accompanying booklets, is a vital aid for professional decision-making and the credibility of the assessment of music performance in the NCEA. The Ministry of Education has provided video exemplars for each level of the music Performance

Standards, with accompanying booklets aimed at assisting teachers in developing their practical knowledge in assessing performance. The title of the Level 1 booklet is “NCEA Level 1 Music Handbook: Guide to Assessing Student Performance”; thus, rather curiously, it avoids use of the word benchmarks or exemplars. The video cassette, however, bears the title “Level 1 Music Exemplars”. That the Standard makes no mention in its explanatory notes of exemplars, benchmarks, or the requirement to use support material provided by the Ministry is interesting, given that it is impossible to make judgements working from the standards document alone. As argued in part one, reference to external factors to anchor the Standard is a fundamental component of the process. This anchoring most commonly involves making an explicit link between teachers’ knowledge of what is likely to be achievable in the field, the Standard, and examples of student work.

Lack of clarity of conception and purpose continues with the support materials provided for performance at Level 2 and 3 of the NCEA. The examples of student work were formatted inconsistently, some issued informally in folders at training days and some in booklet form with videos and cassettes being provided at a later date seemingly carrying a more ‘official’ status. The Level 2 performances are titled “Student Sample” and Level 3 simply bears the title of the Standard. The term exemplar is used inside the Level 2 booklet but not at all at Level 3, and the booklets contain informal suggestions about how they might be used. If the examples of student performances were not intended to act as national exemplars but only as training materials then this was not made clear. Moreover, clarity is not enhanced by the use of the term exemplar on the Ministry of Education resource site Te Kete Ipurangi ([www.tki.org.nz](http://www.tki.org.nz)). Here examples of music assessment tasks are routinely described as ‘exemplars’. In fact, they are task exemplars but do not in any way exemplify a standard as examples of student work are intended to do.

The commentary within the supporting booklets contains some confusing statements in regard to the function the exemplars might fulfil in what is essentially a summative assessment process:

[Y]ou will add your own helpful information as a result of your professional judgement and experience in assessing students against the Achievement Standards. That way it becomes a “living resource” in the hands of the people who are responsible for the teaching and assessment of student learning – those who want the best for students in their care. (Ministry of Education, 2001, p. 3)

The implication appears to be that in order to do the best for students, teachers will incorporate other sorts of knowledge about them as they assess against the Standard, but we are not sure what sort of knowledge this might be. In a summative context teachers should have “an ideological preparedness to engage in making judgements about actual quality” rather than effort or improvement, for example (Sadler, 2005, p. 191). At no point in the booklet is it suggested that the exemplars should be used as a means of anchoring the standard or that they might be utilized as a means of developing a feel for the standard and mitigate against such factors as teachers’ personal expectations, order or serial effects, the halo effect, or comparison with other students, but simply that they are “designed to be a quick

reference for teachers ...” (Ministry of Education, 2001, p. 3). The booklets for Levels 2 and 3 contain similar statements:

[U]se the exemplars as interim benchmarks as [you] begin to accumulate evidence of performances...by [your] own students for use as benchmarks in subsequent years. (Ministry of Education, 2003, p. 1)

Compiling local exemplars is a valid undertaking provided they have been subject to external moderation, but if the suggestion is to replace benchmarks chosen nationally with local ones then this seems counterproductive to a notion of inter-school comparability or a national standard. The Level 3 handbook contains further contradictory statements:

Music is a complex art form and making judgements about an individual’s ability to express themselves both in and through music can be *highly subjective*. It is for this reason that we recommend teachers always refer to the *achievement criteria* within the relevant standard, *as well as drawing on their own experience in making professional judgements and their knowledge of their own students*. (Ministry of Education, 2004, p. 3) (emphasis added)

Just how knowledge of students is expected to inform the summative assessment process is unclear. Presumably ‘objective’ assessments in relation to the criteria can and should be made *without* reference to a student’s circumstances, particularly in a summative assessment context where the possibility of teacher fallibility, such as rewarding effort, for example, needs to be minimised. On the one hand the Level 3 booklet warns against subjectivity, and on the other hand seems to encourage it. Wyatt-Smith and Castleton (2005) and Hawe (2002) report quite distorted assessment practices where teachers were not accustomed to comparing students’ work against standards. While recognising that judgements cannot be entirely objective, we should acknowledge that some procedures encourage more objectivity than others. Knowledge of students is an integral part of formative assessment in a course of study, where teacher feedback is pivotal, but the issue here is the development of judgements for qualifications in relation to published criteria for national standards, and the development of this ability in teachers. In this regard, the booklets appear somewhat contradictory in their intent. The developmental purpose of formative assessment feedback as a regular part of teaching and learning needs to be clearly differentiated from the need at certain times to assess students against a fixed reference point. It can be difficult for teachers to separate these two functions, but the credibility of national standards relies on this. The support booklets do little to clarify these functions.

A further weakness in the booklets is the annotation of points that are not explicitly linked back to the criteria of the Standard. Whatever appears striking or relevant about the performance is noted, and this is very instructive; but it tends to undermine the significance of the actual criteria and reinforce the notion of configurational assessment and latent criteria. Perhaps not unexpectedly the two most subjective criteria - musicianship and presentation - are dealt with the least. In

fact, the Level 1 booklet sets out to deal only with technique and musicianship in any detail (Ministry of Education, 2001, p. 18) and within these sections deals with technique far more than musicianship. This illustrates that certain criteria are easier to deal with through language than others. It is unfortunate that despite the use of videoed performances the discussion of musicianship and presentation was not explored further, and that the criteria were not used as headings to order the annotations provided on each performance.

The linking of teacher feedback to the criteria in the Standard takes on significance in formative teaching contexts when teachers need to identify key features for students at the cut score or threshold level between achievement, merit and excellence. This ability to identify and describe or to ‘unpack’ and make explicit what is required to move student achievement from one level to another is fundamental (Hattie, 2008). The development of this knowledge, initially for a summative function, can also have a positive impact as teachers utilize it in formative feedback with students during production. For example, exemplars developed for the National Curriculum Exemplars Project have encouraged teachers to engage with students in discussions about their work and to focus their expectations for students (Poskitt, Brown, & Taylor, 2003). Exemplars in the senior curriculum can serve a similar purpose by being utilized as tools for teaching and learning (Elley, 1995). Students can develop skills of critical awareness in monitoring the quality of their own work as part of formative assessment processes (Sadler, 1989, 1998).

Much of the training received in the early days of the NCEA focused on teachers devising assessment schedules, but despite this it remains unclear if this is an official requirement (no mention of assessment schedules can be found in the Standard but it is required by NZQA as part of the moderation process). Teachers are encouraged to transform the criteria of the Standard into locally devised tasks with assessment schedules that reflect the assessment activities chosen by the teacher. This process is both a strength and a possible weakness. It allows teachers to shape and emphasize the criteria according to the tasks that they devise, but it makes the process of inter-school moderation more challenging as moderators assess generic qualities in many different tasks. As Pearlman notes, “What they [the content standards] do not do is discriminate among all the qualities and characteristics they articulate ... [t]he assessment design must choose from among all of these qualities ... the assessment represents an interpretation of the Standards document ...” (Pearlman, in press, cited in Moss & Schutz, 2001, p. 54). In the case of performing, which has been the focus for this analysis, the task is clear and the link between the intention of the Standard, any devised schedule, and the practice of music performance is a straightforward one. Such clear and appropriate connections between the assessment criteria and the task might not always be the case in standards with a less practical emphasis.<sup>ix</sup>

## **CONCLUSION - IDEALS FOR PRACTICE**

Those who have criticised the potential lack of rigour in an achievement-based assessment system will find plenty of ammunition in the realities as described

above. However, in an ideal form there seems little reason to believe the system could not deliver sound and credible assessment. The main components are in place, but to ensure rigour and consistency in practice much greater support and clarity of intent is required from the authoritative bodies. The support materials for music were produced by contracted teachers and music advisors who generously sourced authentic student performances (and compositions). Unfortunately, there appeared to be no unified theoretical understanding underpinning the preparation of the materials or the training of the trainers. As a result of this teachers have had to struggle without a clear overview of how an effective standards-based model of assessment should work. Ultimately, government bodies must take responsibility for the lack of clarity and adequate resourcing that have hampered the implementation of major educational change.

National qualifications using a standards-based approach need to be adequately funded to provide systems and on-going training that will give credibility. We should expect standards to appear in a format “that enables them to function unambiguously as the benchmark for valid and reliable assessment ...” (Croft, 1994, p. 18). Standards need to have officially sanctioned explanatory notes and exemplars (linked via the web) that are regularly reviewed and consistent in their intent, presentation and terminology. Teachers should refer to these national exemplars when making judgements. Moreover, ongoing professional development is fundamental to the credibility of the system, not one-off ‘jumbo’ days. As teachers become more comfortable with the practice of standards-based assessment, regular national professional development could provide a mechanism for discussion and focus on wider issues surrounding the NCEA. For example, Locke (2005) and others have highlighted concerns over the fragmentation that can occur in senior programmes under the NCEA. More integrated ways of approaching assessment and task design could be developed as part of continuing discussions.

It has been argued here that standards are not singular stand-alone concepts but come into being through a combination of many factors. Any weak link in the chain (non-utilisation of exemplars, lack of training) has implications for the effectiveness and credibility of the system as a whole and the validity and reliability of assessors’ tasks and assessment interpretations. Shepard (1984) suggests that because “all methods of setting standards are fallible and because there is no ‘true’ standard, information from several sources should be considered ...” (cited in Norcini & Shea, 1997, p. 44). The Ministry of Education is currently undertaking a major review of standards in the New Zealand curriculum. It is to be hoped that this rationalisation and clarification of assessment principles and procedures will be accompanied by the on-going development of exemplar material for all standards, and opportunities for teacher professional development and dialogue. If this were the case, inter-school variations in judgement would be lessened as teachers developed a clearer understanding of the intent and realization of standards-based assessment by participating in a community of practice. Given the limitations of the standards themselves to adequately define a standard, the status of exemplar materials and professional dialogue becomes all the more significant.

## REFERENCES

- Croft, C. (1994). *Standards-based assessment – What do we know? What should we know?* Paper delivered at the 16<sup>th</sup> National Conference of the New Zealand Association for Research in Education, Christchurch, New Zealand.
- Elley, W. (1995). What is wrong with standards-based assessment? In R. Peddie, & B. Tuck (Eds.), *Setting the standard*. (pp.78-98). Palmerston North: Dunmore Press.
- Grudnoff, L. & Tuck, B. (2003). *Professional standards and the professional development of beginning teachers*. Paper presented at the Australian Teacher Education Association Annual Conference, Melbourne, Australia.
- Hattie, J. (2008). *Moderating the moderators*. Presentation for NZQA, Wellington.
- Hawe, E.M. (2002). Assessment in a pre-service teacher education programme: the rhetoric and practice of standards-based assessment. *Asia-Pacific Journal of Teacher Education*, 30(1), 93-106.
- Locke, T. (2005). Talking across the divide: English teachers respond to the NCEA. *Waikato Journal of Education*, 11(2), 113-136.
- Mills, J. (1991). Assessing musical performance musically. *Educational Studies*, 17(2) 173-178.
- Ministry of Education, (2001). NCEA Level 1 Music Handbook: Guide to assessing student performance. Wellington.
- Ministry of Education, (2003). *NCEA Level 2 Music*. Wellington.
- Ministry of Education, (2004). *NCEA Level 3 Practical Music. Booklet to accompany videos and CDs*. Wellington.
- Moss, P. & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38(1), 37-70.
- Moss, P. & Schutz, A. (1999). Risking frankness in educational assessment. *Phi Delta Kappan*, 80, 680-687.
- New Zealand Qualifications Authority (2004). *Achievement Standard Music 2.1 90264*. Retrieved February 2, 2009, from <http://nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2004/as90264.doc>
- New Zealand Qualifications Authority (2005). *Achievement Standard Music 3.7 90776*. Retrieved February 2, 2009, from <http://nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2006/as90776.doc>
- New Zealand Qualifications Authority (2008). *Achievement Standard Music 1.1 90012*. Retrieved February 2, 2009, from <http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2008/as90012.doc>
- Norcini, J. & Shea, J. (1997). The credibility and comparability of standards. *Applied Measurement in Education*, 19(1), 29-59.
- O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, 9(3), 325-335.
- O'Neill, A-M., Clark, J., & Openshaw, R. (Eds.). (2004). *Reshaping culture, knowledge and learning? Policy and content in the New Zealand framework*. Palmerston North: Dunmore Press.



- Peddie, R. & Tuck, B. (Eds.). (1995). *Setting the standards*. Palmerston North: Dunmore Press.
- Poskitt, J., Brown, M., & Taylor, K. (2003). *New Zealand national curriculum exemplars: mist or must for teachers?* Paper presented to NZARE/AARE Conference. Auckland.
- Radnor, H. & Shaw, K. (1995). Developing a collaborative approach to moderation. In H. Torrance (Ed.), *Evaluating authentic assessment*, (pp.124-143). Buckingham: Open University Press.
- Raths, J. (1999). A consumer's guide to teacher standards. *Phi Delta Kappan*, October 1999, 136-142.
- Regelski, T. (2004). *Teaching general music in grades 4-8. A musicianship approach*. New York: Oxford University Press.
- Sadler, R.D. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, R.D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119.
- Sadler, R.D. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, 5(1), 77-84.
- Sadler, R.D. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation*, 30(2), 175-194.
- Sadler, R.D. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education*, 14(3), 387-392.
- Stanley, M., Brooker, R., & Gilbert, R. (2001). Examiner perceptions of using criteria in music performance assessment. *Research studies in music education*, 18, 46-56.
- Thwaites, T. (2001). Desperately seeking standards. *Fanfare, The Journal of the New Zealand Society for Music Education*, 54, 26-37.
- Thwaites, T. (2002). Professor Greyhair's back to the bones of achievement standards moderation. *Music Teach*, 11, 29-30.
- Wyatt-Smith, C., & Castleton, G. (2005). Examining how teachers judge student writing: an Australian case study. *Journal of Curriculum Studies*, 37(2), 131-154.

**APPENDIX 1****ASSESSMENT CRITERIA FOR SOLO PERFORMANCE ACHIEVEMENT STANDARDS NCEA****Level 1 AS90012 Perform contrasting music as a featured soloist**

<b>Achievement</b>	<b>Achievement with Merit</b>	<b>Achievement with Excellence</b>
Perform contrasting music, generally accurately, demonstrating some technical skills, appropriate musicianship and presentation skills.	Perform contrasting music, fluently and mostly accurately, demonstrating technical skills, effective musicianship and presentation skills.	Perform contrasting music, confidently, fluently and accurately, demonstrating secure technical skills, convincing musicianship and communication skills.

(Retrieved February 2, 2009 from <http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2008/as90012.doc>)

**Level 2 AS90264 Present contrasting performances as a featured soloist**

<b>Achievement</b>	<b>Achievement with Merit</b>	<b>Achievement with Excellence</b>
Present generally accurate contrasting performances, demonstrating a range of technical skills, and appropriate musicianship and presentation skills.	Present fluent and mostly accurate contrasting performances, demonstrating a wide range of technical skills, and effective musicianship and presentation skills.	Confidently present fluent and highly accurate contrasting performances, demonstrating secure techniques, and convincing musicianship and presentation skills.

(Retrieved February 2, 2009 from <http://nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2004/as90264.doc>)

**Level 3 AS90776 Prepare and present performances of music as a featured soloist**

<b>Achievement</b>	<b>Achievement with Merit</b>	<b>Achievement with Excellence</b>
Present performances of music as a featured soloist that demonstrate musical understanding and appropriate technical and communication skills.	Present performances of music as a featured soloist that demonstrate musical understanding, secure technical skills and effective communication skills.	Present performances of music as a featured soloist that consistently demonstrate musical understanding, secure and advanced technical skills and convincing communication skills.
Reflect on and evaluate the preparation and delivery of the performances.	Reflect on and evaluate the preparation and delivery of the performances.	Reflect on and evaluate the preparation and delivery of the performances.

(Retrieved February 2, 2009 from <http://nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2006/as90776.doc>)

## APPENDIX 2

### EXPLANATORY NOTES FROM ACHIEVEMENT STANDARD MUSIC 2.1 AS 90264 (VERSION 2)

#### Present contrasting performances as a featured soloist

##### Explanatory Notes

1. This achievement standard is derived from *The Arts in the New Zealand Curriculum*, Learning Media, Ministry of Education, 2000, Level 7 strand, *Communicating and Interpreting Meaning in Music*.
2. *Featured Soloist* means that the student either performs as a soloist with or without an accompaniment, or is easily heard and seen in a small group situation (e.g. piano and another instrument, rock, chamber music, cultural) where the student being assessed plays a separate or uniquely identifiable part.
3. *Contrasting performances* refer to aspects such as style, era, tempo, genre, culture, iwi/tribe, instrument.
4. Performances must be before an audience and may be based in a classroom, school or community. The performances should be video recorded for checking assessments and for moderation purposes.
5. Each performance could comprise a selection of short pieces or an extended piece or pieces.
6. The performances should show evidence of technical skills, accuracy, musicianship and presentation.
  - *Technical skills* refer to techniques specific to the instrument being used for the performance.
  - *Accuracy* refers to the degree of precision as communicated from the written score or the aural transcription, with the understanding that performances are seldom completely accurate.
  - *Musicianship* refers to musical awareness and understanding of the musical style. This includes phrasing, dynamics, rhythm and feel.
  - *Presentation* refers to the sense of performance appropriate to the genre and style of the music. This also includes rapport and communication with the audience, preparation, appropriate posture and stagecraft.
7. The performances should reflect the technical and musical demands equivalent to a fourth year of study through lessons from itinerant teachers.
8. Improvisation skills are assessed where appropriate to the musical style.  
(New Zealand Qualifications Authority, 2004)

---

<sup>i</sup> For discussions on the philosophical, political, and pedagogical aspects of standards-based reforms in New Zealand see O'Neill, Clark, & Openshaw, 2004; Croft, 1994; Elley, 1995.

<sup>ii</sup> In 24 credit courses performance opportunities comprise 9 credits at Level 1, 7 credits at Level 2, and 12 credits at Level 3.

<sup>iii</sup> At Level 1 the equivalent Standard has the title 'Perform contrasting music as a featured soloist' and at Level 3 'Prepare and present performances of music as a featured soloist'. It is unclear if some significance should be inferred from the varied titles, 'performing' music and 'present performances'. In practice it is most likely that the lack of consistency of title is not significant particularly considering the assessment criteria for each level are very similar (see appendix 1). However, such inconsistency in language use is a concern in a high stakes assessment context largely dependent on verbal description.

<sup>iv</sup> The revised Level 1 Standard, AS90012, version 3 (2007) has expanded the equivalent explanatory notes somewhat with the aim of giving further clarification to assist with assessment.

<sup>v</sup> The alternative type of assessment package available in the New Zealand NCEA is a Unit Standard. In this competence-based format the standard is either achieved or not achieved. A single unqualified standard is defined through lists of elements and performance criteria.

<sup>vi</sup> The equivalent statement appears in each Level of this Performance Standard

<sup>vii</sup> The New Zealand Achievement Music Standards were devised by a team of eight writers. The involvement of the writers in training and development of exemplar material has meant a closer connection of those involved in all aspects of the process, avoiding the extreme separation of parts in the process outlined by Moss and Schutz (1999) in the writing of the teacher assessment standards in the U.S.

<sup>viii</sup> For the NCEA, the New Zealand Qualifications Authority (NZQA) has developed moderation procedures that attempt to satisfy the problem of comparability between schools. Schools are required to submit samples of assessed student work each year from a variety of Standards. As well as this, schools are required to have an internal moderation system in place.

<sup>ix</sup> For example, the Music Achievement Standard requiring the study of musical works allows the teacher to choose the works for study and to devise assessment tasks. A very wide variance in type, quality, and difficulty of task is likely as well as the works chosen. No exemplars of student work or detailed assessment schedules have been provided for these Standards.