Peer assessment in musical performance: the development, trial and evaluation of a methodology for the Australian tertiary environment

Ryan Daniel

ryan.daniel@jcu.edu.au

This article outlines the development and trial of peer assessment procedures for implementation within a music performance context in the Australian tertiary environment. An overview of the literature on peer assessment is presented, followed by reference to specific trials of peer assessment within a tertiary music student context. The paper then presents the rationale for and methodology applied in developing two forms of peer assessment of music performance, for trial across one academic year. Student evaluation data obtained via mid- and end-of-year questionnaires are analysed and discussed, as are implications for further research.

Introduction

The awarding of a grade or numerical equivalent to the performance of a musical work can not only be problematic from an external assessor's point of view, but can often be traumatic or frustrating for the student being assessed. Traditionally, the assessment of music performance relies on staff evaluations that most often take place via an end-of-year public performance or recital examination. On many occasions, students' involvement in the assessment of performance is minimal. While external assessment is clearly essential for any student studying musical performance, the issue of how to engage students more actively in the process of assessment requires attention, given the current emphasis in Australian tertiary education on greater student participation in educational procedures.

Some may argue that students at a developmental stage should be subjected to the views of 'experts' only; however, while external assessment has obvious validity, it neglects the development of independent assessment skills. After all, music graduates are required to assess their own performances and the work of others, regardless of whether they move into a performance, teaching, or other career. Employers highlight the need for a range of graduate skills, and as a result there is an increasing emphasis on the development of generic skills, including critical thinking and assessment skills (Brown, 2002; Falchikov, 1998; Brew, 1999).

Considerable attention has been directed towards the value of peer assessment within higher education (Brown, 2002; Falchikov, 1998). At the same time, the use and influence of peer assessment in the music sphere has been less immediate, and particularly so in the Australian context. This may be due in part to the general perception of the process as controversial (Somervell, 1993; Searby & Ewers, 1997; Brown & Glasner, 1999).

Certainly, broadening the assessment processes in music to incorporate the learner is a challenge, although one that has been proved to be very rewarding in other areas of higher education, such as psychology (Falchikov, 1986), history (Ritter, 1998), engineering (Oldfield & MacAlpine, 1995), finance/banking (Pond & ul-Haq, 1998), communication (Weaver & Cotrell, 1986) and medical education (Falchikov, 1995).

In addition to peer assessment practices, self-assessment is gaining increasing attention in tertiary circles (Boud, 1991; Cowan, 1998; Falchikov, 1998; Daniel, 2001). However, it is important to consider that, while each heavily involves the student, these are separate processes (Brew, 1999; Brown, 1998). Peer assessment involves students assessing the work of others, while self-assessment traditionally focuses on a student's self-evaluations. While it is possible to argue that the critical skills required to engage in peer and self-assessment are similar, with some authors referring to the links between the two processes (Somervell, 1993; Falchikov, 1998), they deserve a separate focus, given that each has different requirements.

An overview of the literature

One of the common threads emerging from the literature is the potential for resistance to the implementation of peer assessment, given its move away from the traditional authoritarian approach of assessment by staff only (Searby & Ewers, 1997; Brown & Glasner, 1999; Jordan, 1999). Indeed, Falchikov (1998: 10) argues that involving students in assessment processes is nothing short of 'hard work'. Although peer feedback processes have been introduced in order to create greater efficiencies in providing feedback for students and also for easing pressures on staff (Brown & Glasner, 1999; Lapham & Webster, 1999), the latter has not always been the case, and some authors argue that they create additional workloads and stresses for academics (Brew, 1999; Searby & Ewers, 1997; Hunter & Russ, 1996). Brew (1999: 161), in fact, states that attempting to reduce teacher load and time 'is rarely straightforwardly successful'. Jordan (1999: 172) goes further in suggesting that introducing peer assessment in order to create efficiencies 'is unlikely to work very well if it is artificially grafted on as a time-saving device'.

While it may not necessarily create efficiencies in terms of time and resources, many authors outline the considerable benefits of peer assessment procedures. Weaver & Cotrell (1986), Brew (1999), Jordan (1999) and Lapham & Webster (1999) all refer to the rewards of peer assessment activities when they are well constructed, explained and managed. One of the significant advantages appears to be the manner in which peer assessment assists in developing critical skills: indeed, Brew (1999: 169) argues that 'peer assessment is critical in the development of learning which is genuinely related to the development of life-long learners'. Crooks (1988: 28) also lauds peer assessment, although he argues that it works best when providing a range of feedback 'which [does] not contribute directly to the final grade'. Other perceived benefits indicated include the opportunity for deep as against surface learning (Mindham, 1998), and an improvement in assessment skills, simply as a result of direct engagement in the assessment process (Oldfield & MacAlpine, 1995; Weaver & Cotrell, 1986). It is also noted, however, that some students find it difficult to assess their peers critically (Weaver & Cotrell, 1986; Divaharan & Atputhasamy, 2002).

Contributing factors in the successful integration of peer assessment procedures include adequate preparation and clear explanation of the process in which students are to be

involved (Brew, 1999; Crooks, 1998; Boud, 1995). Nightingale (1996: 124) supports this view and argues that one of the key criteria for the successful introduction of peer assessment is that the process is carefully introduced. This not only involves the design and implementation of, but also the development of criteria for, peer assessment, which should be clear for all parties involved (Nightingale, 1996). Falchikov (1995: 175) presents a similar argument and states that 'the act of making explicit the assessment criteria' is common to successful peer assessment practices.

Peer assessment processes can be divided into two categories, peer marking and peer feedback (criticism). Falchikov (1998) notes that peer marking is designed to require the identification of strengths and suggestions for improvement, with peer feedback presented to enable the student to revise successive drafts of their work, while both offer benefits to students. Brew (1999: 160) favours peer feedback, arguing that peer marking can be problematic and disruptive to students whereas peer feedback can 'contribute to the cohesiveness of student groups'. Peer marking has, however, been applied in several studies which reflect the benefits of this mechanism (Falchikov, 1995; Mindham, 1998; Pond & ul-Haq, 1998), hence it would appear that either process has the potential to be of value for students.

One of the common points of discussion concerning peer and, indeed, self-assessment is the potential for differences between student and staff marks. Many issues arise here and one of the first is the potential for different assessments by staff. Boud (1995: 99) argues that it is 'quite common for different tutors to give the same piece of work widely different assessments', a view also supported by Brew (1999), Falchikov (1986), Simmonds (1988) and Weaver & Cotrell (1986). At the same time, Brew (1999: 161) states that, provided that criteria are in place and are clear, student and staff marks 'tend to be similar'. Falchikov (1995) reaffirms this view and refers to several studies in the medicine and dentistry fields where staff and student assessments of intern performance were similar. One of the main concerns with peer assessment is student over-marking, and Falchikov (1995) proposes that this may be a result of students' hesitance to fail a colleague and/or a lack of knowledge of the area. In general, it would appear that variation in marking has the potential to occur in any situation involving staff and student assessors, although efforts to minimise these aspects should be prominent in the design and presentation of any new procedures.

The timely nature of introducing peer assessment is argued by Nightingale *et al.* (1996: 6), who state that assessment at the tertiary level 'is changing – in its intent and in its methods'. They refer to the three main changes at the tertiary level: the broadening of university education, the need to fully exploit the potentials for assessment and feedback, and the need to involve students in the assessment process so as to enhance the development of a number of skills. Brew (1999) supports the notion of involving students in assessment, and supports a situation where the traditional teacher-centred model of power is shared. Perhaps one of the most important statements made by Brew (1999: 169) is that assessment no longer becomes something done to students but 'an activity done with students', a view supported by Brown (1998).

Case studies – peer assessment practices in music

Searby & Ewers (1997) outline the introduction of peer assessment at Kingston University in order to reduce staff marking time and to speed up the return of feedback. Interestingly,

composition was adopted as the pilot model, on the grounds argument that it was 'less controversial' than musical performance; a potentially problematic view given the complexities of assessing a form where process may be as critical as product. Indeed, Simmonds (1988: 21) argues that there is no greater challenge for staff than 'the prospect of having to assess students' compositions', and refers to a research study where there were notable discrepancies between some staff assessments of student compositions. Searby & Ewers, however, describe an extensive process of educating their students in the new procedures, prior to the trial of group student peer assessment of composition. They argue (1997: 373) that this format of peer assessment did not, in fact, reduce staff time, due to the set-up and facilitation of the process, but that the real benefit was the development of students' critical analysis skills.

Searby & Ewers (1997) then proceed to outline the introduction of a peer assessment system for music performance. Staff and students initially developed assessment criteria, prior to engaging in a number of practice sessions to develop a comprehensive under-standing of the process. Five or six students and a staff member would work together to provide written feedback on performances. Searby & Ewers argue that the process worked well, except for occasional lecturer domination of the process and an insufficient allocation of time to assess students. They raise four concerns in ensuring an effective process, these being

- the method used to arrive at an appropriate mark;
- the provision of adequate written feedback;
- a speedy and efficient system; and
- ensuring the most effective learning environment for students.

In terms of student evaluations of the process, Searby & Ewers found that, as students became more experienced in the process, they found greater value in it. One of the most significant outcomes of the evaluation process was that many students 'felt that staff members on the assessment panels overrode their opinions' (Searby & Ewers, 1997: 379), and the authors note that this requires further exploration.

An additional study into the use of peer assessment in music performance is that outlined by Hunter & Russ (1996). At the time, performance students would perform in public at least once during the course of the year and present an end-of year-examination, these performances being assessed by staff only. Changes were made as a result of considerable student dissatisfaction at the procedures for assessing performance, and of the lack of student understanding of the criteria to be applied. Seminars were introduced where staff and students would engage in informal peer tutoring. Although an improved environment ensued, there was ongoing student dissatisfaction with assessment, so the staff decided to involve students in the assessment procedures. As a result, staff and students then developed the criteria to be applied to performance through a series of panel discussions and trial assessment of performances given by third-year students.

One of the key points argued by Hunter & Russ (1996: 69) is that 'consultation with the students at various stages was vital; we had to develop a sense of student ownership'. Panels of six or seven students were organised to assess performances, with one student assigned the task of preparing a written report. Staff would then meet with the panels to discuss the content of the reports and the performance(s) in question. Hunter & Russ found that, in general, staff marks were 'generally a little lower than those awarded by students'

(1996: 71). Negotiation of marks would then take place where required. Hunter & Russ identified three reasons for unrealistic marks: a reluctance to award low marks, marks reflecting expectations of the performer rather than the performance itself, and marks that were resolved for the sake of expediency rather than after thorough discussion.

The peer assessment processes described by Hunter & Russ (1996) were then expanded to include additional year levels, tutors' reports (which enabled teachers to express their views prior to the performance) and the renegotiation of criteria, with all students being expected to take part in assessment processes. Music was also provided prior to the performance so that students might become familiar with the music, and a standardised report was developed for all performance assessments. The final marks achieved via discussion and negotiation would then contribute to the students' results for that subject. In terms of the benefits of the processes introduced, Hunter & Russ (1996) argue that it encouraged students to prepare more thoroughly for performance; they became more experienced as assessors; and their critical listening, evaluation and negotiation skills improved considerably. In terms of disadvantages, they refer to the doubling of workload for staff, and some student dissatisfaction with having students' assessments contribute to their final marks. Hunter & Russ conclude by arguing that modifications to the process would continue to be necessary.

Synthesis of the literature and directions

The literature reveals that peer assessment, when well constructed, explained and monitored, can be of considerable value to students. Indeed, active engagement in the procedure ensures that students at least *consider* the process of assessment in greater depth than they may have previously. Other potential benefits for students include:

- the development of a range of skills including assessment, critical thinking, evaluation and independent learning skills;
- increased confidence in and knowledge of assessment procedures;
- a greater sense of participation in and control of the learning process; and
- the greater variety of feedback provided.

There are also several aspects that require consideration when introducing such procedures, including the potential for

- initial resistance to and inexperience in such processes;
- an increase in both staff time and resources required;
- some students' reticence to assess their colleagues' work critically;
- character conflicts and personality problems for some students when assessing as members of a group;
- student dissatisfaction where peer assessment contributes to the final grade; and
- variation between student and staff assessments.

It is also clear that additional research in the area of student evaluation data is necessary, a view supported by Falchikov (1995: 177), who argues that few student evaluations of peer assessment are reported. Although Brown, Sambell & McDowell (1998) present student views on the introduction of peer assessment, views which support the positive aspects of such procedures, these are generic and originate from a large sample of data gathered

over four years of peer assessment trials. Similarly, the qualitative comments provided by Weaver & Cotrell (1986) and Falchikov (1986) are not only limited, but are specific to fields other than music. In terms of the music case studies presented above, student evaluation data is scant. Hunter & Russ (1996) fail to present any student evaluation data. Searby & Ewers (1997) provide statistical data, and although these present a mostly positive view, qualitative comments are absent. For instance, Searby & Ewers (1997) report that the majority of students felt that changes to the processes implemented were required; however, written detail of these proposed changes was either not requested or not presented.

Thus the current research will focus on two main areas:

- the opportunity to involve students in structured peer assessment procedures and which do not necessarily impact directly on weighted assessment; and
- the need to explore further and present detailed student perceptions and evaluations related to the process within a music performance context.

Background

Prior to the commencement of the 2002 academic year, music performance students at James Cook University engaged in a weekly performance seminar during which students performed in order to gain verbal staff and student feedback. Given that no formal requirement to perform was in place, more motivated students performed frequently, while others would attempt to delay performance, or in some cases avoid it completely. Similarly, student feedback often derived from the same body of students. Although prompted by staff, many students appeared to find the process of verbalising performance feedback overly challenging. Nor did students have any form of written feedback as a basis for engaging in self-reflective processes. The system appeared flawed, and as a result, a feedback mechanism was developed in order to consider how to enhance the system prior to re-implementation in 2002. A number of potential scenarios for operating the class were presented, and students ($n = 36$) were required to rate these in order of preference and to propose suggestions for creating the most effective learning environment. The feedback generated from the short questionnaire revealed that the majority (78%) felt that

- all students should be required to perform on a regular (rostered) basis; and
- a structured system of both staff and student feedback should be implemented.

The remaining students (22%) indicated that a more structured system of performance should be implemented, although with less structured feedback and analysis of performance processes. All students therefore indicated that the system required greater structuring, and as a result, significant changes were made for the 2002 academic year.

Methodology

All students were scheduled to perform twice per semester, so four times per academic year. Rather than allocate a weighted assessment component to these performances, it was decided to view them as formative assessments, in order to enable students to gain experience performing a number of works and to obtain a range of feedback prior to

the end-of-year exam. To ensure that students would approach the formative assessments seriously, students were required to complete the four performances satisfactorily in order to be eligible for the final examination, which formed the major assessment component of the performance subject at each year level. The first step, therefore, was to incorporate staff assessments, and it was arranged that two academics would evaluate all student performances using the criteria sheets which students were accustomed to from end-of-year examinations. These criteria sheets required the detailed assessment of musical literacy, technical skills, interpretation and professional skills, corresponding to the standard university system of high distinction, distinction, credit, pass and unsatisfactory. It was decided that, firstly, it would be inappropriate at this early stage to require students to assess peers using the extensive two-page criteria sheets used by staff, and secondly, it would also be impractical to require students to assess all performances. Rather, it was deemed most important to ensure that the peer assessment mechanisms considered and/or developed would take into account the students' general lack of experience, while at the same time directly involving them in the process of assessment of performance, albeit in a less involved manner at this stage.

The next step was the development of appropriate peer feedback mechanisms, with students involved in the design process. Given the novelty of the procedures to be introduced, and indeed the students' relative inexperience in participating in assessment design to date, staff prepared two potential models of peer assessment for the purposes of obtaining student feedback and discussion. The objectives, construction and student requirements for the two proposed models are outlined in Table 1.

Table 1 *Peer assessment mechanisms presented to students for feedback*

Mechanism	Objective	Construction
General peer appraisal	To provide students with a short evaluative overview of their performance.	Student assessors to present three short evaluative comments regarding the performance.
Detailed peer marking and feedback	To provide students with detailed diagnostic feedback and a grade (mark).	Students to grade a range of areas using five-point scales, present an overall grade with relevant justifications, and indicate three strengths and three weaknesses.

During the first five weeks of the semester, a range of activities and discussions took place in order to consider the potential peer assessment procedures. Students were initially informed of the rationale for, purpose and potential benefits of peer assessment. Additional issues discussed were the relevance and validity of peer criticism and feedback, and it was acknowledged that although the introduction of peer assessment would in fact create more work for staff and students, the potential benefits were significant. Students were provided with the two mechanisms for peer assessment, and a number of case study examples (using previously videotaped performances) enabled students to try out the methods to consider their suitability.

The introductory sessions proceeded relatively smoothly and student reactions were mostly positive, although some negativity was encountered as a result of the increased workload and the requirement to engage in new and arguably challenging procedures. The case study practicals did, however, prove to be particularly useful for students, in that they were able to directly participate in and gain an understanding of the function and purpose of the peer assessment mechanisms. Debate ensued as to which of the methods would be most suitable, and given that many students felt that each would be of value, it was agreed that both mechanisms be adopted. At the same time, some revisions were made to the structure of the sheets. For instance, students argued for the inclusion of 'fluency of performance' as an area to be evaluated on the detailed marking sheet. In addition, the requirement to identify up to 'three weaknesses' on this same sheet was changed to 'three areas requiring the most attention', in order to present a more positive tone for those receiving the assessments. While some students raised concerns regarding potential problems with the procedures, such as negative and/or irrelevant feedback, it was acknowledged by staff and students that it would be necessary to trial the procedures for a continuous period in order to consider the success or otherwise of the new approach.

Interestingly, considerable debate took place regarding the number of assessments that students should be required to complete, as well as the choice of performer. Some argued the necessity for complete freedom in the choice and number of peer assessments, while others (including staff) presented the view that some students might be favoured, some receive few or even no assessments, and some invest minimal effort in the procedures. Ultimately, it was agreed that students should be provided with one detailed peer marking sheet and one sheet containing three short appraisals, requiring them to assess 4 of the 10–12 performances scheduled per week. The choice of performer for the detailed marking sheet was left to the student, while it was determined that in order to enable all students to gain at least some feedback, the general appraisals would be numbered in a format to ensure that all performances were assessed. The general appraisal sheets were subsequently designed as follows, and an equal number of each sheet were to be disseminated throughout the student body:

- Sheet 1: peers to assess performers 1, 5 and 9;
- Sheet 2: peers to assess performers 2, 6 and 10;
- Sheet 3: peers to assess performers 3, 7 and 11; and
- Sheet 4: peers to assess performers 4, 8 and 12.

On those few occasions where there were less than 12 performances, those students with sheets 3 and 4 were to choose an alternate performer to assess where appropriate. The full version of the detailed marking peer assessment mechanism and an example (sheet 1) of the general appraisal sheets are provided as Appendix A.

The two methods of peer assessment, in addition to staff evaluations, would form a three-way feedback mechanism for students. The performance coordinator would collect all feedback sheets at the end of the class, review them to ensure that there were no untoward comments, and have them available for collection as soon as possible after the performance class. In addition, it was acknowledged that detailed feedback would be obtained at the end of the semester in order to investigate students' reactions to the procedures over time and

in the context of 'live' peer assessment. The peer assessment methods were subsequently implemented in week 6. Initial observations at the end of the semester (week 13) were that:

- the new procedures promoted a considerably more structured approach to performance practice;
- students evidenced greater need to prepare adequately for performances; and
- all students were now engaged in the process of critical assessment of performance.

While these observations were valid, it was essential to ascertain the views of the entire student cohort in a more structured and formalised manner, prior to re-implementation in semester 2. A short two-page questionnaire was developed to enable students to evaluate each of the peer assessment processes from the perspective of both performer and assessor, to consider changes in their ability to assess performance, and to suggest enhancements to existing procedures. All performance majors completed the questionnaire, which at the time included 16 first-year, 6 second-year, and 12 third-year students ($n = 34$).

In addition to implementing a questionnaire at the end of semester 1, a second questionnaire was developed for application at the end of semester 2. This questionnaire was designed to gain additional feedback from students regarding the procedures continued in semester 2, to evaluate changes made, and to consider the extent to which students' skills and perceptions related to peer assessment and evaluation had been enhanced across the year. Some questions were the same as those in the first questionnaire in order to allow a direct comparison across semesters. All performance majors completed the questionnaire at the end of semester 2, which at the time involved 31 students. One student had withdrawn from the course and two others did not return the questionnaire, leaving a sample of 12 first years, 5 second years and 14 third years.

Questionnaire data

Procedural aspects

At the end of semester 1, students were initially asked to present an overall evaluation of the value of each of the peer assessment mechanisms (see Table 2). The detail provided in the marking sheets evidently led to students perceiving greater value in this form of peer assessment. In order to investigate this further, students were firstly required to rate the value of each process for them as a performer receiving feedback, and to indicate the relevant reasons for their choice. Table 3 presents these data for semesters 1 and 2, in order to be able to consider any significant change across the academic year.

Table 2 *Evaluation of overall value of the peer assessment mechanisms*

Detailed marking sheet	General appraisal sheet

Mean grading
  (1 low, 3 moderate, 5 high)
	Mean grading
	  (1 low, 3 moderate, 5 high)

4 (SD 0.87)	2.47 (SD 0.92)

Table 3 *Students' preferred feedback sheet as* performer

Form of peer assessment	End S1	Identified reasons (summarised, no. of comments) $n = 34$	End S2	Identified reasons (summarised, no. of comments) $n = 31$
Detailed marking sheet	85%	• Additional detail and/or information provided (18) • More thought involved • Greater volume of comments • Opportunity to consider identified strengths and weaknesses	87%	• More information, detail and feedback (23) • More area specific (3) • Appraisal sheet comments can be misinterpreted
General appraisal sheet	9%	• Quicker to read • 'Bluntly helpful'	3%	• Sometimes people tend to 'fill up' the marking sheets
Both	6%	• Both sheets provide useful feedback (first year) • Allows for comparison between both styles of feedback (third year)	10%	• Both give valuable feedback • Marking sheets provide detail and feedback sheets give overall view (2)

The data in Table 3 clearly indicate the majority of students' preference for the detailed marking sheet, given the additional information and range of specified feedback. Statistically insignificant change is observed across semesters. At each semester, a small number of students preferred either the general appraisal sheet or a combination of the two, and the change across the year was minimal. Students were also required to indicate their preferred method for assessing peers; these responses are presented in Table 4.

Table 4 reveals that most students prefer the detailed marking sheets when assessing peers. What is interesting, however, is that at both questionnaire points, some students prefer to use the general appraisal sheet when assessing others, although their first preference as a performer is to receive the detailed marking sheet. This may be due to the additional workload and time involved in completing the detailed marking sheets. A common-sense view is that raised by a third-year student at the end of semester 1, supporting the value of both sheets depending on the specialist area of the performer to be assessed. The second challenging statement is by a third-year student at the end of semester 2, who argues that neither process is sufficient for them. This may be influenced by the fact that experience at these processes has led to a more critical view on the opportunity for maximising the benefits of these two peer assessment methods.

Critical assessment issues

In order to consider the degree of students' critical assessment, respondents were asked to indicate the extent to which they felt peers were appropriately critical when evaluating performances at the end of the first semester, and to consider whether this had improved or developed by the end of the second semester (Table 5).

Table 4 *Students' preferred feedback sheet as* assessor

Form of peer assessment	End S1	Identified reasons (summarised, no. of comments) *n* = 34	End S2	Identified reasons (summarised, no. of comments) *n* = 31
Detailed marking sheet	68%	• Additional detail and/or information provided (11) • Easier to assess given detailed requirements for completion • More thought involved – 'thoroughness' • More depth and thought required	74%	• More specific, hence easier (8) • More accurate and/or detailed assessment (10) • Provides more information (3) • Forces me to think of positive aspects as well
General appraisal sheet	20%	• Thinking of descriptive words *really* makes me think • Gives more of a general overview • Easier to fill in • Allows more time to listen to the performance	13%	• You can be more critical even when you don't understand the instrument • It's easier just looking for three aspects (2)
Both sheets	12%	• Both allow you to write what you feel is necessary • It is good to have both available • The detailed sheets are best for same instrument assessments while general feedback is good for other instruments (third year)	10%	• Marking sheets are good when wanting to provide detail, while feedback sheets are good for general or quick assessments (3)
Neither	0%	N/a	3%	• When marking, I would prefer the score, while with feedback, I need to explain in more detail (third year)

Table 5 *Students' perceptions of the degree of peers' critical assessments*

Mean – Semester 1 (*n* = 34) *(1 not critical enough, 4 appropriately critical, 7 over-critical)*	Mean – Semester 2 (*n* = 31) *(1 much less critical, 4 about the same, 7 much more critical)*
2.97 (SD 1.65)	4.79 (SD 1.17)

At the end of the first semester, students felt that their colleagues were not appropriately critical, which, given the novelty of these procedures, is not unexpected. What is valuable, however, is that most students view peer assessments to have become more critical by the end of the year. While staff made it known to students at the commencement of semester 2 that the majority of students felt peers were not sufficiently critical, students identified a range of reasons for an improvement in the degree of critical assessment, including:

• students are getting used to it;
• assessments are more critical, extensive and accurate;

- students over time have improved and become used to the criteria;
- students have a better understanding of instruments and techniques now;
- students listen for different things now; and
- assessments have become more in-depth as people understand how useful they are as a performer.

The last comment is particularly interesting and reveals the fact that students rely on their peers to some extent to present valid assessments. The data therefore indicate that experience of the process, in addition to the cajoling of students by staff to produce relevant assessments, led to some improvement in students' critical assessment skills over the year. While this may not be accurate across the entire student cohort, it supports the need to continue to refine these processes in order to ensure that peer feedback is as valid as possible. Students were also asked to compare student assessments of performance with those of staff at the end of each semester (Table 6).

Table 6 *Comparison of student assessments with staff evaluations*

Year level Semester 1 ($n = 34$) Semester 2 ($n = 31$)	Much more critical than staff		About the same as staff		Much less critical than staff	
	Sem. 1	Sem. 2	Sem. 1	Sem. 2	Sem. 1	Sem. 2
First years (no. of students)	0	1	6	9	10	2
Second years (no. of students)	1	0	0	0	5	5
Third years (no. of students)	0	1	1	2	11	11
TOTAL no. of students	1	2	7	11	26	18

Table 6 reveals that the majority of students considered peer assessments during the year to be much less critical than those of staff. This may be due to the students' hesitation to be critical, their relative inexperience at performance assessment, or their lack of knowledge of the area, or staff may indeed be overly critical. At the same time, there is a slight change by the end of semester 2, in that some students felt that peers had become more critical when assessing performances, so validating the results presented in Table 5. Certainly, these data also suggest that further work in the area of improving the validity of peers' critical assessment is required, or that the issue of staff and peer assessments should be further explored to examine the extent to which staff and/or students are appropriately critical.

Semester 2 enhancements

In order to consider the requirements for completing the peer assessment mechanism, students were asked to rate the workload and difficulty of each at the end of semester 1. The responses are presented in Table 7.

The data reveal that for the majority of students, the requirements were manageable, particularly so in relation to the general appraisal sheet; however, students were asked to suggest what revisions might be made to existing procedures. Five first years and six third years, or 32% of the total group, felt no changes were necessary. Other suggestions are summarised in Table 8.

Table 7 *Evaluation of workload and difficulty of peer assessment mechanisms*

Aspect evaluated (*n* = 34)	Detailed marking sheet Mean grading *(1 low, 3 moderate, 5 high)*	General appraisal sheet Mean grading *(1 low, 3 moderate, 5 high)*
Workload	3.06 (SD 1.05)	1.91 (SD 0.86)
Difficulty	2.94 (SD 1.11)	2.06 (SD 0.84)

Table 8 *Summary of students' suggestions regarding semester 2 procedures*

Year level	Suggested enhancements
1	• Use detailed sheets only (4) • Make general sheets more detailed • Only same instrument assessments • General feedback/discussion after the performances
2	• Detailed sheets only and more time to complete these • Drop general sheets and include discussion after the performances • Include an open forum on what students and staff felt about the performances • Everyone should have at least three detailed sheets per performance – no general feedback • No changes other than a minimum of two detailed sheets per performer
3	• Discuss points after performances • More detail is needed on the general feedback sheets • People should be encouraged to be more honest • Remove general sheets or make them short answer evaluations • Use students who take the time to consider and assess accurately

The responses presented in Table 8 reveal the following issues in relation to student opinions:

• there should be greater flexibility in the choice of feedback sheets;
• students should be more critical; and
• there is a need for an open discussion of the performances at the end of the class.

While several students argued that the general appraisal sheets should be discontinued, some students had previously indicated a preference for these sheets as assessors and performers, hence the complete removal of this option would potentially disadvantage some students. In addition, the fact that 32% of students indicated that no changes were necessary suggests that one-third of the group found value in continuing existing

procedures. However, revisions were considered based on the emerging principles from the student feedback, with the proposed amendments and relevant objectives presented in Table 9.

Table 9 *Analysis of revisions regarding semester 2 classes*

Emerging principles from feedback	Proposed amendment	Objective
Greater flexibility in choice of feedback sheets.	Student choice of both the type of sheet and the number of assessments.	To enable greater freedom in choice of assessment mechanism and to allow students to assess same instrument only.
More critical assessment by students.	Additional workshops with students regarding appropriateness of language used and degree of critical assessment.	To further enhance students' ability to critically assess performance and to provide performers with more valid feedback.
Lack of interaction and discussion following performances.	To include at the completion of the class a time for open discussion and interaction between students and staff.	To enable performers to receive verbal feedback and assessors to engage in an open and interactive discussion of performances presented.

At the commencement of semester 2, detail of the feedback regarding potential changes was presented to students, and two sessions were devoted to the use of critical language in the assessment area and its validity for performers. For instance, those students who used terms such as 'good' or 'great' were advised that a more diagnostic explanation would assist performers. The new system was discussed and agreed upon, with students given a choice of feedback sheets and a short open forum for discussion incorporated at the end of the class. At the end of semester 2, students were asked to indicate the degree to which the greater flexibility in the choice of feedback sheets was appealing. Using a seven-point scale, from *not at all appealing* (1) to *extremely appealing* (7), the mean of 5.29 and standard deviation of 1.37 suggests that the flexibility was mostly advantageous for students. Common reasons identified and which support the flexibility included the following:

- Flexibility enabled choice of marking sheets for own instrument and feedback for others;
- It enables me to use the sheets that I feel are most beneficial;
- I can choose the sheet I feel most comfortable using; and
- I can choose how much effort to put into assessing others.

Students were then required to evaluate the value of the open discussions at the end of classes, using a seven-point scale from *not at all valuable* (1) to *extremely valuable* (7). The mean of 5.23 and standard deviation of 1.36 suggests that most students found at least

moderate value in the open discussion. When asked to express written views about this new process, responses were various and are summarised in Table 10.

Table 10 *Summary of students' views on the open discussion process*

Year level	Positive comments	Negative comments	Neutral comments
1	• It's nice to know what people think. (4) • Good to hear comments quickly. (4) • Good comments make me feel good. • People who don't write can add comments in this forum.	Unfair when 'elite' performances are praised and others forgotten.	Still find out how you performed via the feedback sheet.
2	• Very valuable to all performers. • Immediate feedback. • Allows further comments/discussion.	• Students are unable to provide useful oral feedback. • Verbalising that you performed badly is not good.	*None*
3	• Students are more involved. • The whole group can hear feedback. • It is extra feedback. • The feedback is fresh. • You learn from others' comments. • Students are more verbally active. • It is easier to talk than to write.	• Could be more in-depth and not just generalised statements. • Students don't understand my area hence comments are not valuable.	Most comments are generalities.

The advantages for students are considerable, although some students appeared to find difficulties in this environment. The emerging principles from the negative feedback are that comments need to be specific and relevant, and also that there is the potential for unpleasant experiences for those students either not mentioned or who are presented with negative comments. Indeed, in terms of both the written and the verbal mechanisms, some students continue to comment on the fact that many comments are either invalid or inappropriately critical for them. While this is problematic, it is arguably the case that some students will always be hesitant or unable to present appropriately critical assessment, while this may also be influenced by the fact that the tasks did not contribute directly to their assessment. Responses to suggested changes or enhancements to semester 2 procedures are presented in Table 11.

The majority of students (71%) felt no changes were necessary, which reveals a greater degree of satisfaction with the procedures as compared with semester 1. The revisions to allow greater flexibility in the choice of feedback and the inclusion of opportunities for open discussion proved to be of value to the processes. At the same time, it is envisaged that the following strategies will further enhance those procedures adopted in semester 2:

• greater student involvement in the verbal discussion and training in valid and reliable assessments;
• encompassing all performers and performances in the discussion; and
• greater staff involvement in terms of assisting students to improve their feedback skills and for such activities as feedback on feedback.

Table 11 *Suggested enhancements to semester 2 procedures*

Year level	Suggestions
1	• No changes needed (9) • None – somehow make more people talk • Maybe a private talk would be beneficial and just as good as a group talk • Each person should have to make a comment, so no one is left out
2	• None (3) • Perhaps comments after each performance • Not forcing us to complete all aspects of the sheet
3	• None (10) • Maybe more discussion on a one-to-one basis • More marking sheets to be used to increase feedback • Randomly choose individuals to read out their comments and critique their comments to improve students' assessment skills • More staff feedback

Overall evaluations

Table 12 presents students' views on the advantages and disadvantages of the peer assessment procedures introduced.

Table 12 *Summary of identified advantages and disadvantages of new procedures*

Year level	Advantages	Disadvantages
1	• More feedback/views made possible • Makes us think • Allows you to know where you are at • Understanding improves • Understand the audience reaction • Creates an awareness of your flaws • Identifies areas to work on	• Some students don't make an effort • Not as helpful as the lecturers' comments • Sometimes comments are very critical • Varying grades received • Some are definitely too nice • Comments that are not at all helpful • No feedback on feedback
2	• Broad scope of opinions • Variety of feedback • I now have to listen to and think about each performance • Gives us experience at marking	• Students' minimal knowledge of other instruments • Sometimes you receive vague comments
3	• It is interesting to see the difference between staff and student feedback • Interesting observations that staff may miss or not pick up on • Honest feedback • More opinions and more directions • Good training at assessment	• People are often not honest (critical) • Differences of opinion • Frustrating to be constantly told about a problem you are working hard at correcting • It makes the performers nervous • More pressure and stress! • Comments might not be relevant

The responses reveal that there is a range of both advantages and potential disadvantages for students. The advantages presented are not only numerous but support the notion that regardless of the potential for variation in peer assessments, it is the additional feedback provided that is a major benefit for students. In terms of disadvantages identified, some of these relate to individuals, such as a lack of effort at critical assessment, vague comments, lack of detailed knowledge of other instruments, or the additional stress involved for performers. As such, these appear unavoidable. It is potentially the case, however, that they may diminish as students become more experienced at the process, and especially given the fact that these procedures had only been in place for one academic year.

Students were also requested to present an overall evaluation of the procedures for each semester, in order to consider the extent to which changes made in semester 2 were beneficial (Table 13).

Table 13 *Student ratings of the overall value of the peer assessment processes*

Mean – Semester 1 ($n = 34$)	Mean – Semester 2 ($n = 31$)
(1 not at all beneficial, 4 moderately beneficial, 7 extremely beneficial)	*(1 not at all beneficial, 4 moderately beneficial, 7 extremely beneficial)*
4.44 (SD 1.14)	5.16 (SD 0.32)

While the result for semester 1 reveals moderate satisfaction, it is pleasing that by the end of the second semester, students perceive the processes to be more valuable. While the semester 2 rating does not reveal an overwhelming change in students' perceived value, the movement is quite positive and suggests that with ongoing refinements, the procedures may in fact offer increasing benefits for students over time.

Benefits for students

Students were required to consider the extent to which they felt that their ability at assessing performance had improved since the introduction of the peer assessment processes. The responses are summarised in Table 14.

It is pleasing that the majority (91%) of students felt they had improved in their ability to critically assess performance, a key statistic in supporting the notion that although not without difficulties, the benefits of peer assessment processes are identifiable and wide ranging. Students were then asked to identify the reasons why they felt they had improved in this area (Table 15).

The responses in Table 15 reveal that, amongst other factors, exposure to a range of performances and engagement with peer assessment practices are two of the main contributors to the development of students' critical skills.

Table 14 *Students' indication as to improvements in their ability to assess performance*

Year level	Comments indicating improvement, with explanations (summarised)	Comments indicating undecided or lack of improvement, with explanations
1	Percentage of total comments – 41% *I've realised what was helpful and what was not.* *I'm a lot more confident and understand what to do.* *I feel that I have improved greatly.* *I've learnt new words and ways of assessing.*	Percentage of total comments – 6% *I can't really tell.* *I'm not sure why but I feel that my mindset of thinking critically hasn't really been brought to the fore or been reawakened yet. Maybe this is me!*
2	Percentage of total comments – 18% *I feel I have improved significantly.* *I find it easier to assess performances now.* *I am becoming more confident with my comments.*	*No comments*
3	Percentage of total comments – 32% *Much improvement.* *I feel more qualified to comment than previously.* *I am learning more about performances on other instruments.* *I feel my abilities have improved greatly.*	Percentage of total comments – 3% *Not really.*

Table 15 *Students' indication as to the reasons for the improvement in their assessment skills*

Year level	Example comments indicating reasons for improvement
1	*Because making us do this is good practice.* *Experience, because it is a weekly thing and you get lots of practice.* *I've learnt to listen for things that I didn't when I first started.* *Being able to see other students' comments has been very helpful.*
2	*I've heard more performances on different instruments.* *Practice at the process.* *Due to the continual practice at assessing students every Tuesday.* *The addition of peer assessment sheets.*
3	*The more performances I hear the easier it is to discern between good and bad performances.* *Continued persistence from staff!* *I feel I have more knowledge of musical performance so I feel more qualified to comment.* *From listening to many performers.*

Summary and implications

The research project reveals that the introduction of structured peer assessment procedures has a considerable impact on students' critical abilities in assessing performance in both

written and verbal formats, and that a range of benefits are presented to those students who are awarded the various assessments. While it is clear that the procedures implemented in this study require ongoing revisiting and enhancements, the evidence obtained via questionnaires suggests that students gain and continue to develop a number of skills by simply engaging in the process of critically assessing performance. While some may approach peer assessment in a casual or even negative manner, and as a result some students continue to refer to insufficiently critical evaluations, the majority appear to view the potential benefits of peer assessment as sufficient motivation to contribute in a positive and constructive manner. Certainly, the perceived advantages identified outweigh the disadvantages, hence continuation of the processes is not only valid but arguably essential for music performance students at this level. While there is an additional increase in workload for both the subject coordinator and the students, it is this additional workload that leads to the provision of a more holistic performance feedback environment for students, and the opportunity to extend confidence and skills in critical assessment.

One of the first implications of this study is the need for ongoing refinements to the processes in place, given that some disadvantages have been identified, most notably concerning the perceived lack of validity of general appraisal assessments. It may therefore be necessary for staff to consider how to create a more detailed model, yet one that continues to allow for several short assessments within the one performance session, given the need to ensure that all students receive some feedback. A number of additional implications emerge from this study. Firstly, there is a need to analyse student and staff assessments over time, not only to compare these assessments and consider relevant influences, but to examine the extent to which experience at assessment is a direct contributor to the development of students' critical assessment skills. Secondly, it is worth considering to what extent students' performances improve as a result of engaging in peer assessment. Indeed, case study analyses and/or comparisons of individual student evaluations and performances over time may reveal insights into the parallel development of critical assessment and performance skills. Thirdly, future trials might implement weighted assessment of students' peer assessments in either verbal or written forms, in order to determine whether this will encourage students to engage with the process more diligently. Fourthly, the implementation of focus sessions where students prepare for assessments via study of the relevant performance score may contribute to the validity of peer feedback. While there is considerable room to explore revisions to the processes defined above, and indeed new innovations within the area, students who engage in the study of music performance are advantaged when taking part in structured peer assessment activities.

References

BOUD, D. (1991) *Implementing Student Self-Assessment*. Cambelltown, NSW: Higher Education Research and Development Society of Australasia.
BOUD, D. (1995) *Enhancing Learning through Self Assessment*. London: Kogan Page.
BREW, A. (1999) 'Towards autonomous assessment: using self-assessment and peer assessment'. In Sally Brown & Angela Glasner (Eds), *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, pp. 159–71. Buckingham, UK: Open University Press.
BROWN, A. (2002) 'Peer assessment in tertiary music: a summary of the issues'. Report presented to the National Council for the Heads of Tertiary Music, Australia.

BROWN, S. (1998) 'Introduction'. In Sally Brown (Ed.), *Peer Assessment in Practice*, pp. 5–7. Birmingham, UK: SEDA Publications.

BROWN, S. & GLASNER, A. (Eds) (1999) *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Buckingham, UK: Open University Press.

BROWN, S., SAMBELL, K. & MCDOWELL, L. (1998) 'What do students think about assessment?' In Sally Brown (Ed.), *Peer Assessment in Practice*, pp. 107–12. Birmingham, UK: SEDA Publications.

COWAN, J. (1998) *On Becoming an Innovative University Teacher*. Buckingham, UK: Open University Press.

CROOKS, T. J. (1988) *Assessing Student Performance*. University of New South Wales, NSW: Higher Education Research and Development Society.

DANIEL, R. (2001) 'Self-assessment in performance'. *British Journal of Music Education*, **18**, 3, 215–26.

DIVAHARAN, S. & ATPUTHASAMY, L. (2002) 'An attempt to enhance the quality of cooperative learning through peer assessment'. *Journal of Educational Enquiry*, **3**, 2, 72–83.

FALCHIKOV, N. (1986) 'Product comparisons and process benefits of collaborative peer group and self assessments'. *Assessment and Evaluation in Higher Education*, **11**, 2, 146–65.

FALCHIKOV, N. (1995) 'Peer feedback marking: developing peer assessment'. *Innovations in Education and Training International*, **32**, 2, 175–87.

FALCHIKOV, N. (1998) 'Involving students in feedback and assessment'. In Sally Brown (Ed.), *Peer Assessment in Practice*, pp. 9–21. Birmingham, UK: SEDA Publications.

HUNTER, D. & RUSS, M. (1996) 'Peer assessment in performance studies'. *British Journal of Music Education*, **13**, 67–78.

JORDAN, S. (1999) 'Self-assessment and peer assessment'. In S. Brown & A. Glasner (Eds), *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, pp.172–82. Buckingham, UK: Open University Press.

LAPHAM, A. & WEBSTER, R. (1999) 'Peer assessment of undergraduate seminar presentations: motivations, reflection and future directions'. In S. Brown & A. Glasner (Eds), *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, pp. 183–90. Buckingham, UK: Open University Press.

MINDHAM, C. (1998) 'Peer assessment: report of a project involving group presentations and assessment by peers'. In S. Brown (Ed.), *Peer Assessment in Practice*, pp. 45–65. Birmingham, UK: SEDA Publications.

NIGHTINGALE, P. (1996) 'Module 5: accessing and managing information'. In P. Nightingale *et al.*, *Assessing Learning in Universities*, pp. 115–26. Sydney, NSW: University of New South Wales Press.

NIGHTINGALE, P., TE WIATA, I., TOOHEY, S., RYAN, G., HUGHES, C. & MAGIN, D. (1996) *Assessing Learning in Universities*. Sydney, NSW: University of New South Wales Press.

OLDFIELD, K. A. & MACALPINE, J. M. K. (1995) 'Peer and self-assessment at tertiary level'. *Assessment & Evaluation in Higher Education*, **20**, 1, 125–32.

POND, K. & UL-HAQ, R. (1998) 'Assessing using peer review'. In Sally Brown (Ed.), *Peer Assessment in Practice*, pp. 23–44. Birmingham, UK: SEDA Publications.

RITTER, L. (1998) 'Peer assessment: lessons and pitfalls'. In S. Brown (Ed.), *Peer Assessment in Practice*, pp. 79–85. Birmingham, UK: SEDA Publications.

SEARBY, M. & EWERS, T. (1997) 'An evaluation of the use of peer assessment in higher education: a case study in the school of music, Kingston University'. *Assessment & Evaluation in Higher Education*, **22**, 4, 371–83.

SIMMONDS, R. (1988) 'An experiment in the assessment of composition'. *British Journal of Music Education*, **5**, 1, 21–34.

SOMERVELL, H. (1993) 'Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment'. *Assessment & Evaluation in Higher Education*, **18**, 3, 221–33.

WEAVER, R. L. & COTRELL, H. W. (1986) 'Peer evaluation: a case study'. *Innovative Higher Education*, **11**, 1, 25–39.

APPENDIX A: PEER ASSESSMENT MECHANISMS

1. Detailed peer marking sheet

Name of performer: _____ Accompanist (if any): _____

Name of work performed: _____

Please CIRCLE the X on the scale that most closely relates to your impression of the performer's presentation:

PERFORMANCE ASPECT	POOR		AVERAGE		EXCELLENT
Accuracy of notation	X	X	X	X	X
Dynamic variation	X	X	X	X	X
Tone colour variation	X	X	X	X	X
Technical control	X	X	X	X	X
Fluency of performance	X	X	X	X	X
Stylistic interpretation	X	X	X	X	X
Professionalism	X	X	X	X	X
Ensemble work, if applicable	X	X	X	X	X
Overall presentation	X	X	X	X	X

What grade would you award to this performance? _____

(Choose one from High Distinction, Distinction, Credit, Pass, Unsatisfactory)

Please explain why you chose the above grade -- give at least two reasons:

_____
_____
_____

What were the three major strengths of this performance?

1 _____
2 _____
3 _____

What areas, in your opinion, need the most attention?

1 _____
2 _____
3 _____

PLEASE USE THE REVERSE PAGE OF THIS SHEET TO ADD ANY ADDITIONAL COMMENTS YOU FEEL MAY BE VALUABLE FOR THE PERSON(S) THAT YOU HAVE ASSESSED.

2. Peer assessment general appraisal (sheet 1)

Name of Performer 1: _____ Instrument: _____

Provide three short evaluative comments, which in your opinion, most accurately describe/evaluate this performance (please note that adjectives such as *good*, *ok*, *fine*, or *not bad* are unacceptable):

1 _____
2 _____
3 _____

Name of Performer 5: _____ Instrument: _____

Provide three short evaluative comments, which in your opinion, most accurately describe/evaluate this performance (please note that adjectives such as *good*, *ok*, *fine*, or *not bad* are unacceptable):

1 _____
2 _____
3 _____

Name of Performer 9: _____ Instrument: _____

Provide three short evaluative comments, which in your opinion, most accurately describe/evaluate this performance (please note that adjectives such as *good*, *ok*, *fine*, or *not bad* are unacceptable):

1 _____
2 _____
3 _____