

Performance Assessment: Lessons from Performers

Kelly A. Parkes
Virginia Tech

The performing arts studio is a highly complex learning setting, and assessing student outcomes relative to reliable and valid standards has presented challenges to this teaching and learning method. Building from the general international higher education literature, this article illustrates details, processes, and solutions, drawing on performance assessment and studio research literature in the field of one-to-one, or applied, music teaching. The nature of musical performance assessment will be explained as an emerging tool embedded in the pedagogical methodology research literature of the applied studio learning setting. Implications are explored and suggestions are made for other disciplines in regards to both summative and formative performance assessment strategies.

Assessment in higher education has been under scrutiny since 1990 (James & Fleming, 2004-2005, p. 51), and music assessment can be included in this area “identified by those in the measurement community as prime examples of unreliable measurement” (Guskey, 2006, p. 1). It is now recognized that assessment provides a critical link in the teaching and learning process and that in higher education, researchers are exploring assessment techniques in a variety of discipline settings. This paper will examine assessment’s role in higher education in the 21st century and its place in the wider literature. Case studies from various disciplines will be illustrated and the focus will move to performance assessment and how performance arts faculty develop and use tools to evaluate literal ‘performance.’ Analogous relationships may be seen with other disciplines that require ‘performance’ be assessed, or evaluated, and suggestions are made for developing tools to measure student performance.

Assessment in Higher Education

“Few educators receive any formal training in assigning marks to students’ work or in grading students’ performance and achievement” (Guskey, 2006, p. 2), and yet the importance of higher education teachers’ understanding of the fundamental underpinnings of the principles of assessment can be seen in some of the most recent literature (Brown, 2004-2005; Guskey, 2006; James & Fleming, 2004-2005; Orell, 2006; Stefani, 2004-05; Van den Berg, Admirall, & Pilot, 2006). Shepard (2000) puts forward a sound framework regarding the importance of assessment in learning cultures, using public school classrooms as her platform for discussion, and she makes several good points that can be applied directly to higher education. The first point states that assessment should be seated in the middle of the teaching and learning process (p. 10) instead of being postponed to the end-point of instruction. Shepard calls

this “dynamic” assessment and points out that these assessments are usually found in teaching and learning settings. Her second point poses that feedback, as part of assessment, should not only consist of reporting right or wrong answers to students but that scaffolding and expert tutoring techniques are more successful. She cites the work of Lepper, Drake, and O’Donnell-Johnson (Shepard, 2000, p. 11), who found that tutors often ignored student errors when they were not important to the solution, and they prevented students from making errors a second time by gently hinting or asking leading questions. This type of indirect feedback was shown to maintain student motivation and self confidence as it was used throughout the learning episode. Shepard’s third important point refers to transfer, where she suggests that we assess our students’ abilities to “draw on old understandings in new ways,” and she adds the notion that assessment should not merely test “familiar and well-rehearsed problems” (p. 11). Not surprisingly, this leads to her fourth point regarding the explicitness of the criteria in assessments. The clarity and specificity that music as a performance discipline has had to adopt in the criteria for assessments has been a long uphill struggle and will be expounded upon later in this article.

Shepard speaks also of transparency in good assessments (p. 11), which expresses to students the characteristics of excellent performance. Noticeably, this also assists students in assessing themselves and each other. Shepard’s fifth notion points to the promise that student self-assessment holds for “increasing students’ responsibility for their own learning and to make the relationship between teachers and students more collaborative” (p. 12). These five elements (dynamic or ongoing, feedback, transfer, criteria, and self assessment) form the basis of the positive characteristics of assessment and the remainder of this section will discuss how higher education in general is examining them and then the final section will illustrate how the performing arts utilize them.

Stefani (2004-05) posits that faculty in higher education in New Zealand need to understand the fundamental principles of assessment and maintains that assessment is an integral part of student learning. She acknowledges that higher education now requires more than just “transmission of knowledge” (p. 51) and that academic faculty now must design, develop, and deliver accessible curriculum to an ever-diversifying body of students. She stresses the importance of “teaching, learning, and assessment” (p. 53) and, like Shepard (2000), would like to see assessment embedded within the teaching and learning setting, rather than just “added on after course content has been decided upon” (Stefani, 2004-05, p. 54). Stefani also puts forward key processes and, again like Shepard (2000), setting criteria, sharing responsibility between faculty and students, ensuring transparency of assessment criteria, and providing useful feedback (Stefani, p. 63-64).

Orell (2006) takes a ‘snapshot’ approach to examining feedback as it is used in Australian higher education academic practice, labeling it “the cornerstone of all learning, formal and informal” (2006, p. 441). Like Stefani (2004-05), she notes “the provision of formative feedback as an *add-on* to teaching and learning responsibilities” (p. 442). She argues, however, that “providing students with focused, comprehensive feedback on their learning product is a significant aspect of teaching and assessing” (Orell, p. 442-443). She, along with Shepard (2000) and Stefani (2004-05), present the notion that feedback can affect students’ construction of themselves, motivate future learning, and affect how faculty see the student-teacher relationship. Her study examined how academic faculty in teacher education and nursing fields gave feedback, what it indicated to students, and what kind of congruence there was between the feedback actually given and what the faculty viewed as feedback.

Brown’s (2004-2005) findings similarly support the idea that the ways in which we assess our students can affect how they learn. Her ‘fit-for-purpose’ (2004-2005, p. 81) argument offers insight about the use of “portfolios, in-tray exercises, posters, annotated bibliographies, reflective commentaries, critical incident accounts, reviews, role-plays, and case studies” (p. 83) as alternative methods of assessment seen in the United Kingdom setting. She presents the most important tenets of assessment to be efficiency, transparency, inclusivity, and reliability, and suggests that feedback is the principal area. Brown’s position supports that of Shepard (2000), Stefani (2004-05), and Orell (2006).

Van den Berg, et al. (2006) focus on peer assessment. They report that feedback is adequate in the higher education setting in the Netherlands when it is used formatively and summatively for products as well

as performed in small groups. Their work focused largely on written feedback, which seemed to be more successful when delivered orally as part of a discussion with the reviewer. The afore mentioned researchers (Brown, 2004-2005; Guskey, 2006, April; Orell, 2006; Stefani, 2004-05; Van den Berg, et al., 2006) agree on the importance of feedback as part of the teaching and learning process. Feedback is also one of Shepard’s (2000) five important areas, and the remaining four; dynamic or ongoing assessment, transfer, criteria, and self-assessment, can also be seen in the literature that illustrates the diverse impact these strategies have on student learning in the higher education setting.

Impact on Student Learning in Higher Education

O’Donovan, Price, and Rust (2004) bring forward the importance of students’ understanding of criteria in the assessment standards within higher education in the United Kingdom. They note, “the secret nature of assessment deliberations is no longer seen as acceptable” (p. 326). This point has had wide implications within the performing arts, specifically music, and will be addressed in the next section of this paper. Calvert (2004-2005) reported that developing and using grade descriptors for all tutors in the media communications field of the University of Gloucestershire, UK, was met with mixed results. Tutors, who are faculty that provide smaller sessions for students in between large weekly lectures, found it difficult to agree on descriptors in a one-day session. They created holistic rubric for all tutors to use in grading written work, but reports from tutors revealed they simply ‘adapted’ the descriptors to their own existing grading processes. Student responses were also mixed with comments indicating they preferred more personalized comments added to the rubric as part of the feedback. This mixed response to changes in assessment practice has been seen in the music performance research literature also (Parkes, 2007; 2010).

O’Donovan, et al. (2004) also found many obstacles in making assessment criteria transparent, such as the clear articulation of criteria, the different levels of expectation within the criteria, and the many interpretations that were made with simple terminology such as “synthesis or analysis” (p. 327). Their initial concern was with having experts agree so that the students or novices could also understand the assessments. Like Shepard (2000), O’Donovan, et al. spoke about the inherent importance of transferring knowledge so that students have an awareness of explicit transfer processes. These typically include giving students “explicit learning outcomes, marking [grading] criteria, and eventual feedback” (2004, p. 331). They add that other transfer processes are also

effective, such as “dialogue, observation, practice and imitation to share tacit understandings”(p. 332). These types of transfer can be seen with much more frequency in the performing arts assessment modus operandi of music. Ultimately, and regardless of discipline, transfer does require feedback as part of the process.

Gibbs and Simpson (2004-2005) proposed a tiered approach to feedback with several ‘conditions’ whereby assessment can impact student learning. The ten conditions largely relate feedback to the assessment tasks themselves. Conditions 1 and 2 stipulate that the number of assessment tasks is appropriate to the amount of study time and that the tasks allow students to allocate the correct amount of effort to the aspects of the course (p. 12-14). Conditions 3 and 4 suggest that the assessment tasks engage students in productive learning activity and that feedback is given on the tasks often and in suitable detail (p. 14-17). Conditions 5, 6, 7, and 8 speak to the focus of the feedback: that it is on the work, rather than the student himself; that the focus is kept on the task in time for the student to receive further help; that the feedback is appropriate to the task; and that the feedback is appropriate in relation to what the students think they are supposed to be doing (p. 18-21). Conditions 9 and 10 require that students receive and attend to feedback and that they act on it (p. 23-24).

The work of Macdonald (2004-2005) in the United Kingdom offers an alternative method of grading physics papers and moves the grading process to the students, who evaluate their own papers by using a ‘self evaluation document’ devised by the lecturer. The author reported difficulties in convincing students that it was a fair and defensible practice, but she did find that 80% of students graded themselves appropriately. Zoller (2004-2005) also examined students’ self-assessment of homework assignments in a higher education organic chemistry course in Israel. Findings were promising, as students showed appropriate grading of themselves in line with the professor’s final scores. Peer assessment is often not warmly embraced by students, as Connor (2004-2005) points out in her study findings from the United Kingdom. She asked students in health profession courses to undertake ‘inter-professional workshops’ (p. 98), which required students to work together in groups to foster presentation, collaboration, evaluation, and personal contribution skills. Students created a portfolio and gave a presentation of the work they learned, and they also completed essays discussing key issues in inter-professional collaborations. Students reported skepticism initially about the process, fearing inequity or imbalances in the group. However, Connor reports that after the process and course were completed, students and examiners evaluated the learning process and assessments as “valuable” (p. 101). Presentations as assessments were found to be successful by Brothers

(2004-2005) in students enrolled in counseling programs in the United Kingdom. The presentation skills themselves were not assessed, only the content for its relevance and links to practical application and practice. Feedback was issued in small groups, and students were also asked to write a personal reflection to show their growth over time. Students reported that this type of experience was “the most powerful learning experience” (p. 91).

Nestal, Kneebone, and Kidd (2004-2005) explored scenario-based assessments of technical skill in undergraduate medical education in the United Kingdom. Simulated models were linked with actors so that students could develop clinical skills in a real work setting. The key elements of their scenario-based assessments were ‘preparation, performance of procedure, reflection, and feedback’ (p. 108). Semi-structured interviews, after the assessments were carried out, revealed that students found the process constructive but that peer-evaluating each other was less helpful because they perceived all of their knowledge to be at the same limited level. Robinson and Udall (2004-2005) of the Southampton Institute in the United Kingdom examined the impact of instructor-led conversations about the quality of learning outcomes as part of an assessment strategy to encourage learners to initiate conversations about their own learning in engineering. Students were asked to participate in sessions to make self-assessments of whether they were meeting course outcomes. Students recorded their progress and noted questions they had for the tutor about outcomes they had not met. Findings from this action research showed that students understood why they were being asked to complete certain assessments and gained a heightened sense of understanding about their progress.

Jenkins (2004-2005) proposed that computer-aided assessment can be a largely motivating process, particularly in feedback, as it is accessible online more frequently and comprehensively. Jenkins examined the use of Information and Communication Technology (p. 68) in the United Kingdom and found that it can be used for diagnostic, formative, and summative assessments. He reported advantages as being “repeatability, ... reliability, diversity, timeliness, ... motivating to students, and being student-centered” (p. 68). He puts forward a variety of forms for these assessments such as multiple-choice tests, case studies, online portfolios, personal reflections on weblogs, online mock exams, audits, and group discussions on weblogs. He cautions interested readers about the challenge of culture change, from the UK perspective, to embrace online learning, but stresses that “online formative assessment produced benefits in terms of flexibility and immediacy of feedback” (p. 78).

James and Fleming (2004-2005) also point out that the key features of assessment in the United Kingdom higher education system include feedback and motivation. Their study illustrates the various methods of assessment such as ‘report writing, essay writing, poster presentations, and oral presentations’ (p. 44) used within programs of study and that students don’t perform consistently better on one form of assessment than any another. It is interesting to note this study’s innovative examination of agreement in student performance through the testing of traditional assessments, which is in contrast to Jenkins’ (2004-2005). The variety with which these different disciplines approach the many areas of assessment is commendable and congruent with how the performing arts trajectory for assessment development has also occurred. It is of interest then to also examine another type of assessment, more directly relevant, such as authentic, or to use the more objective term, *performance assessments* (for a detailed history of the choice of the term performance assessments over authentic assessments, please consult Newmann, Brant & Wiggins (1998) and Terwilliger (1997)).

Performance assessment has received interest within the educational literature for several decades, and in some ways this indicates the relatedness to the arts. The arts are a performance discipline, and most assessments there within are concerned with an actual performance itself. In terms of assessment at large, this has now become a term for a type of assessment, one that is different to a standardized assessment or a cognitive test. Concerns have traditionally, within the assessment, measurement, and evaluation literature, been focused on validity and criteria and aligning measurement standards. Early work by Linn et al. (Linn, Baker, & Dunbar, 1991) set forward the notion that criteria, amongst other issues, are the most salient in complex, performance-based assessment, albeit in public school settings, but conceptually the considerations remain the same. Performance assessments must have transparent criteria, be fair, be generalizable and transferable, have cognitive complexity and content quality, and be comprehensive. These are often expected from standardized assessments and should also be apparent in performance-based assessments. Linn (1994, p. 9) went on to elaborate regarding the difficulties that face performance-based assessment, namely reliability and validity. He states that when care is taken in training raters to use well-defined rubrics, reliabilities improve (p. 10). He suggests that to accurately assess students’ achievement, more than one task may need to be assessed (p. 10). Linn’s further work with Swanson et al. (Swanson, Norman, & Linn, 1995) holds up the health profession as a detailed example of performance-based assessment model, and there are noticeable

similarities between these strategies and what is found in the music performance assessments: the tests are conducted in realistic performance situations, but that there are still discrepancies between this situation and real-life; scoring can be problematic; and a selection of assessment methods should depend on the skills to be assessed (Swanson, et al., 1995, pp. 6-8, 11). Delandshere & Petrosky (1998) examine the meaningfulness and usefulness of numerical ratings for the assessment of complex performances, and while their findings refer to the performances of teachers, similar problems present themselves in music performance assessment. Delandshere & Petrosky asked the judges in their study to assign numerical ratings and to draw inferences. This occurs in music performance assessment where numerical ratings allow evidence of reliability to be calculated, and inferential information often appears as a global grade accompanying an overarching set of comments in the music performance research literature, as noted by Bergee (2003).

Music Performance Assessments

The above research and theoretical frameworks illustrate how important assessment—in particular ongoing assessment, feedback, transparency of criteria, and self-assessment—is in the higher education teaching and learning setting. Generally speaking, these above disciplines have found methods and strategies to be successful in the way they impact student learning, and this impact occurs in a positive way not only for the student but also for the faculty, as it informs them for future teaching. Disciplines that are outside the scope of traditional lecture-based, or even small group tutor-led teaching have had to create their own methods of assessment, and this can be seen for example in the health profession. Traditional multiple-choice bubble-tests, standard essay, or even written assignment models do not apply as an appropriate assessment tool to test student achievement on the content being taught. The strategies that the arts, such as dance, theatre, and specifically music, have employed for assessment have always been ‘performance’ based, both as type of assessment and as a literal explanation of the process. The applied music studio, where students learn individually, in the music conservatorium has come late to the ‘assessment movement,’ perhaps because, according to Schleuter, “good, bad, and inefficient methods and techniques [including assessment] of teaching music persist though unquestioned adherence to tradition” (Schleuter, 1997, p. 20). The Western music conservatoire has been in the business of ‘conserving tradition’ for hundreds of years. Performance disciplines such as music are usually taught via the master-apprentice model, whereby the student comes to learn once a week in a very focused,

complex environment, one-to-one with the master teacher, typically a recognized performing artist. The nature of formative assessment has, in the past, been seen in feedback throughout the session. It refers to technique, musicianship, and stylistic features. The formal, summative assessment typically occurred at the end of the semester in the form of a performance, called a jury, and was graded by an unspoken global system between expert faculty judges. Judges are applied performance teachers and are experienced performers of, and listeners to, music at very high levels. Remarkably, these end-of-semester jury performance exams have ranked high in inter-rater reliability (Bergee, 2003). In some cases, criteria have been debated between judges, often heatedly, as to their importance at this final performance stage. After discussion between the judges, the final grade was given to the student on a form with some comments about improvements or achievements. In some parts of the United States this *modus operandi* still exists, yet in others it is markedly different. The applied music studio, and the use of assessment within, has only been investigated and thus reported on over the past 20 years. It is with optimism that a new wave of research emerges highlighting inclusion of assessment as part of the teaching and learning process.

Examining the work that some of the leading music performance assessment researchers have conducted, a similar trend can be seen in strategies, but interestingly there is sparse connection in literature reviews to other disciplines. The music education researchers pioneering this line of inquiry are moving in a similar vein to much of the literature discussed in this paper, especially in the areas outlined by Shepard (2000), namely dynamic and ongoing assessments, feedback, transfer, criteria, and self-assessment. The current paper will proceed to illustrate some of this research and offer practical implications for developing performance assessment tools.

Historical Perspectives

The applied music studio has been embedded in Western art music since its earliest settings, and most will recognize this teaching and learning setting as the 'master-apprentice' model. Almost all musicians in Western Art music genre have learned their musical skills through this traditional method, and by speaking about Western art music, popular music and informal learning are excluded because those teaching and learning settings are usually markedly different. Assessment is part of the teaching and learning setting and the development of skills. Colwell (1971) suggested that "it is often thought that music teachers are against systematic evaluation because they fear the exposure of poor teaching" (p. 41), and he added that a

more important reason might be the "conservatory atmosphere." He states that "conservatories are trade schools; their emphasis is on the development of specific skills ... these skills are constantly evaluated ... lessons, recitals ... are constantly filled with evaluation" (p. 41). Applied faculty are making assessments and evaluations, and have been doing so for many years, particularly in the 'conservatory atmosphere,' so there is much to be gained from examining the research literature for insight into how this is being done.

Mills (1987) has suggested that in the Western art music tradition of assessment, a vocabulary is used in the discussion of performance for the purposes of evaluation, and some applied faculty prefer the verbal openness of the comment sheet at recital or jury time to convey feedback about a music performance. There is a long held oral tradition in the applied studio, and the vocabulary is often instrument-specific. The work of Duke (1999), however, has identified teacher feedback as a specifically useful tool used in lessons given in the Suzuki school method. The term *feedback* is usually used in music education literature to refer to instant reinforcement that occurs within short teaching frames, whereas the term assessment is generally used synonymously with grading or summative evaluation. Feedback in applied music lessons occurs with much higher frequency, similar to the tutor model of feedback that Shepard (2000, p.11) cites from Lepper, Drake, and O'Donnel.

Duke & Simmons (2006) reveal that musical goals and expectations are prominent elements in lessons given by internationally renowned artist-teachers. The connection between these goals and expectations and the assessment points for measurement or evaluation are not made by Duke & Simmons; however, it is clear that the expectation of the artist-teacher is that the student play in a lesson as if they are performing on stage in order to achieve "a high standard" (p. 12). This type of feedback is conveyed to the student consistently, and it is reasonable to assume the student knows that this expectation continues to prevail in the jury or recital setting. The jury or recital setting is predominantly the authentic or performance assessment, in the truest sense of the word 'performance.'

The works of Bergee (2003) tested the reliability and validity of specific criteria rating scales, or rubrics, in the college applied studio setting in an attempt to create tools that would be reliable and valid for the summative assessment of musical performances. His findings support the concept that the criteria help the applied music faculty grade more consistently in the jury setting, and Bergee also showed that they grade with more reliability if they use a tool with specific criteria as opposed to giving a 'global' grade based on an overall impression of the performance. The use of a

Table 1
Brass Criteria Specific Performance Rubric

Assessment Categories	Acceptable 14-15	Proficient 16-18	Exceptional 19-20	
Interpretation / Expression (Includes dynamics)	<ul style="list-style-type: none"> Has acceptable stylistic qualities Makes some attempts to play with stylistic appropriateness Plays with a reasonable musical effect Make regular attempts at pleasing phrasing 	<ul style="list-style-type: none"> Has proficient stylistic qualities Usually plays with stylistic appropriateness Plays with proficient musical effect Consistently uses pleasing phrasing 	<ul style="list-style-type: none"> Exceptionally stylistic Always plays with stylistic appropriateness Plays with exceptional musical effect Always uses the most pleasing phrasing 	— 20
Tone	<ul style="list-style-type: none"> Tone, in general is acceptable Tone is mainly consistent across registers 	<ul style="list-style-type: none"> Tone, in general, is clear Tone is maintained proficiently across registers 	<ul style="list-style-type: none"> Tone, in general, is extremely clear Tone is exceptional across all registers 	— 20
Intonation	<ul style="list-style-type: none"> Intonation is adequate but is inconsistent some of the time within the player and / or accompaniment Demonstrates some understanding of tonality 	<ul style="list-style-type: none"> Intonation is proficient and only small inconsistencies appear within the player and / or with accompaniment Demonstrates proficient understanding of tonality 	<ul style="list-style-type: none"> Intonation is exceptional and no inconsistencies appear within the player / and or accompaniment Demonstrates exceptional understanding of tonality 	— 20
Technique	<ul style="list-style-type: none"> Shows acceptable posture Holds instrument with competence Plays correct notes (fingering and / or pitching) Has acceptable specific technical skills – transposition, clefs, mute changes, hand-stopping Shows minimal problems with embouchure 	<ul style="list-style-type: none"> Shows good posture Holds instrument with confidence Plays correct notes with confidence (fingering and / or pitching) Has proficient specific technical skills – transposition, clefs, mute changes, hand-stopping Shows no problems with embouchure 	<ul style="list-style-type: none"> Shows great posture Holds instrument with bravura Plays correct notes all the time with exceptional confidence Has exceptional specific technical skills – transposition, clefs, mute changes, hand-stopping Shows strong embouchure 	— 20
Rhythm / Tempo	<ul style="list-style-type: none"> Short periods of consistent tempo Melodic rhythm approximately correct Tempo changes sometimes observed from music Sometimes rushing/dragging 	<ul style="list-style-type: none"> Consistent tempo most of the time Melodic rhythm consistently correct Tempo changes always observed from music No disruptive rushing/dragging 	<ul style="list-style-type: none"> Tempo was consistent all the time Melodic rhythm precise all of the time Tempo changes always observed with exceptional skill Never rushing/dragging 	— 20
TOTAL				

specific tool in the applied studio measurement process is innovative yet has not been embraced by many conservatories or music departments. The criteria rubrics are typically analytic rubrics where the elements of the musical performance are identified individually, with descriptive statements across a continuum of scoring illustrating the levels of attainment. See Table 1 for an example.

Parkes (2007) tested the use of a criteria-specific performance rubric with applied faculty (*n* = 5). She tested both student and faculty attitudes towards assessment prior to the use of a criteria specific performance rubric in lessons and juries. She then post-tested students and faculty to attempt to ascertain the perceived benefits to either students or faculty. She found some resistance from the faculty in using the tool, to some degree because they did not perceive a

need to use a measurement instrument in an otherwise unchanged protocol of jury comment-sheet grading. Her rubrics, for brass and woodwind instruments, both yielded internal consistencies of .97 and .93 respectively, but her findings in regard to student and faculty perceptions about the use of the rubric were not significant due to low participation. The later work of Parkes (2010) found that when used for self-assessment, a criteria specific performance rubric assists students in a more learner-centered approach to their improvement. One applied faculty professor offered to use the performance rubrics with her students during lessons, and she also asked students to use the rubric to evaluate their own performances, which were recorded in each lesson. The professor additionally asked the students to reflect each week in an online journal about their playing and what they heard while

evaluating themselves with the rubric. Findings from this pilot case study suggest that students had positive perceptions about the rubric, that their awareness about their own improvement was increased, and they had a clear understanding of what was required by their professor each week. These findings support the notions of Shepard (2000) in regards to transparency of assessments and increasing students' responsibility for their own learning.

Oberlander (2000) didn't examine the use of specific measurement tools in the applied studio, but she did investigate the grading procedures in general. She showed that the overwhelming majority of clarinet instructors in the Northern USA and Canada give grades for applied studio learning based on effort and improvement. Oberlander recommended that a fixed criterion be used in determining final grades to gain a higher level of objectivity, which supports the works of Bergee and Parkes. Oberlander suggested determining in advance what level should be reached in order to pass particular criteria; keeping a written record of each lesson, and possibly assigning a grade for each lesson; and having final grading techniques involve a screen to maintain anonymity of the students. The findings of Oberlander support Shepard's (2000) commitment to transparency of criteria in assessments and contribute to the overall dialogue about what could improve assessments in the applied studio.

The work of Ciorba and Smith (2009) was initially conducted in response to the recent push from accreditation bodies in requiring the implementation of specific assessment tools. In this study, a multidimensional assessment rubric was administered to all students performing a jury recital ($n = 359$). The results of this study indicate that there was a high degree of inter-judge reliability where reliability coefficients were above .70, which is not surprising. However, of more interest is the process by which the rubric was developed. It was not solely developed by the researchers, as in the work of Bergee and Parkes, but by a panel of faculty who, over the course of one semester, identified common dimensions shared across all areas and created descriptors outlining the various levels of achievement. The panel then piloted the rubrics over the following two semesters to refine the rubric and the practicality of its use in a jury or final performance setting. The rubric was used across all instrument and voice areas, and the findings reported that performance achievement was positively related with participants' year in school, with a one-way multivariate analysis of variance. This indicates that when a faculty group get invested and involved with what they are looking for in student achievement, they can create assessment tools that meet their needs. In line with Shepard (2000), this study exemplifies the importance of providing feedback for students and

making the criteria clear as to the characteristics of excellent performance.

Conclusion

The features of assessment as explained by Shepard (2000) can be seen in the higher education literature across several countries and, more importantly, the research of music performance literature. The practical solutions that performance arts such as music have adopted are remarkably similar to Shepard's notions of what constitutes effective assessments. By attempting to use performance assessment rubrics within lessons and at the end of semester, Parkes (2010) illustrates the potential for this type of ongoing feedback for students. By using criteria-based feedback, both Bergee (2003) and Ciorba & Smith (2009) make the case for improving faculty specificity and improving student performance. Oberlander's (2000) suggestions also highlight the need for clarity in criteria. It is important to recognize these findings and seat them in the higher education literature as they represent valid and reliable ways to measure what can sometimes be seen as a 'subjective' discipline. Faculty who participated in the music performance research studies had to outline for themselves what excellent performance should look like and then bring these expectations to their colleagues and students. The notion of transparency in assessment is required for assessments to capture student achievement reliably. Music performance assessments are moving away from "the secret nature" of past practice that has been criticized by many (Brand, 1992; Jones, 1975; Madsen, 1988, 2004; Schleuter, 1997). By examining and defining the required components of music performance, researchers in this area have been able to move this discipline forward, creating the critical link between teaching and learning in more applied music settings. The concept of true 'performance assessment' is seen in this subset of research, and it is garnering some momentum in the ways applied music performance faculty are teaching and their students are learning. This move is a positive one and perhaps one from which other disciplines may benefit. Such benefits might start with promoting good communication between faculty, fostering a desire to remove the secretive or subjective nature of assessments, developing a willingness to embrace new methodologies, and ultimately testing and refining their effectiveness in the teaching and learning setting.

References

- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137-150.

- Brand, M. (1992). Voodoo and the applied music studio. *The Quarterly Journal of Music Teaching and Learning*, 3(2), 3-4.
- Brothers, D. (2005). A presentation: A preferred route for assessment in counsellor training. *Learning and Teaching in Higher Education*, 1(1), 90-92.
- Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education*, 1(1), 81-89.
- Calvert, B. (2005). Demystifying marking: Reflections on developing and using grade descriptors. *Learning and Teaching in Higher Education*, 1(1), 93-97.
- Ciorba, C. R., & Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric. *Journal of Research in Music Education*, 57(1), 5-15.
- Colwell, R. (1971). Musical achievement: Difficulties and directions in evaluation. *Music Educators Journal*, 57(8), 34-43, 79-83.
- Connor, C. (2004-2005). An aligned assessment to promote learning about collaboration between health and care professionals. *Learning and Teaching in Higher Education*, 1(1), 98-101.
- Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27(2), 14-24.
- Duke, R. A. (1999). Teacher and student behavior in Suzuki string lessons: Results from the international research symposium on talent education. *Journal of Research in Music Education*, 47, 293-307.
- Duke, R. A., & Simmons, A. L. (2006). The nature of expertise: narrative descriptions of 19 common elements observed in the lessons of three renowned artist-teachers. *Bulletin for the Council of Research in Music Education*, 170, 7-19.
- Gibbs, G., & Simpson, C. (2004-2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1(1), 3-30.
- Guskey, T. R. (2006). "It wasn't fair"- Educators' recollections of their experiences as students with grading. Paper presented at the American Educational Research Association, San Francisco, CA. Retrieved from http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICEExtSearch_SearchValue_0=ED492005&ERICEExtSearch_SearchType_0=no&accno=ED492005.
- James, D., & Fleming, S. (2005). Agreement in student performance in assessment. *Learning and Teaching in Higher Education*, 1(1), 32-50.
- Jenkins, M. (2004-2005). Unfulfilled promise: Formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education*, 1(1), 67-80.
- Jones, W. J. (1975). Games studio teachers play. *Music Journal*, 33, 46.
- Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher*, 23(9), 4-14.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Macdonald, A. (2005). Student self-evaluation of coursework assignments: A route to a better perception of quality. *Learning and Teaching in Higher Education*, 1(1), 102-107.
- Madsen, C. K. (1988). Senior Researcher Award Acceptance Speech. *Journal of Research in Music Education*, 36(3), 133-139.
- Madsen, C. K. (2004). A 30-year follow-up study of actual applied music versus estimated practice. *Journal of Research in Music Education*, 52, 77-88.
- Mills, J. (1987). Assessment of solo musical performance: A preliminary study. *Bulletin for the Council of Research in Music Education*, 91, 119-125.
- Nestel, D., Kneebone, R., & Kidd, J. (2004-2005). Triadic peer review in scenario-based assessments. *Learning and Teaching in Higher Education*, 1(1), 108-111.
- Newmann, F., Brandt, R., & Wiggins, G. (1998). An exchange of views on "semantics, psychometrics, and assessment reform: A close look at 'authentic' assessments". *Educational Researcher*, 27(6), 19-22.
- 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.
- Oberlander, L. M. (2000). *A survey of procedures in applied clarinet at American and Canadian colleges and universities*. Unpublished doctoral dissertation, Arizona State University, Tempe, AZ.
- Orell, J. (2006). Feedback on learning achievement: Rhetoric and reality. *Teaching in Higher Education*, 11(4), 441-456.
- Parkes, K. A. (2007). Effect of performance rubrics on college level applied studio faculty and student attitudes. In T. Brophy (Ed.), *Integrating Curriculum, Theory, and Practice: A Symposium on Assessment in Music Education* (pp. 325-333) Chicago, IL: GIA Publications.
- Parkes, K. A. (2010). The use of criteria specific performance rubrics for student self-assessment: A case study. In T. Brophy (Ed.), *The Practice of Assessment in Music Education: Frameworks, models, and designs*. (pp.453-458). Chicago, IL: GIA Publications.
- Robinson, A., & Udall, M. (2004-2005). A framework for formative assessment: Initiating quality learning

- outcomes. *Learning and Teaching in Higher Education*, 1(1), 112-115.
- Schleuter, S. L. (1997). *A sound approach to teaching instrumentalists*. Belmont, CA: Schirmer.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Stefani, L. (2004-05). Assessment of student learning: promoting a scholarly approach. *Learning and Teaching in Higher Education*, 1(1), 51-66.
- Swanson, D. B., Norman, G. R., & Linn, R. L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher*, 24(5), 5-35.
- Terwilliger, J. (1997). Semantics, Psychometrics, and Assessment Reform: A Close Look at "Authentic" Assessments. *Educational Researcher*, 26(8), 24-27.
- Van den Berg, I., Admirall, W., & Pilot, A. (2006). Designing student peer assessment in higher education: analysis of written and oral feedback. *Teaching in Higher Education*, 11(2), 135-147.
- Zoller, U. (2004-2005). Supporting 'HOCS learning' via students' self-assessment of homework assignments and examinations. *Learning and Teaching in Higher Education*, 1(1), 116-118.
-

KELLY A. PARKES, a professor of music education in the Department of Teaching and Learning at Virginia Tech, has focused her research efforts on the pedagogy of teaching pre-service music educators. She focuses on the motivation and self-efficacy of pre-service teachers, and the impact of these on teachers' performance. She has also made efforts to increase the use of research-based assessment strategies in higher education music education settings. Dr. Parkes has published in journals such as *Update: Applications of Research in Music Education*, *Journal of Music Teacher Education*, *Journal of Research in Music Performance*, *Journal of the College Music Society*, *the International Journal of Teaching and Learning in Higher Education*, and the *Bulletin of the Council for Research in Music Education*.