

The Journal of Research in Music Performance (JRMP)

[Archives](#)

[Editorial Board](#)

[Submissions](#)

Recent Research in Applied Studio Instruction: Evaluation and Assessment

By Kelly A. Parkes

The purpose of this paper, the third in what has become a three-part series, is to examine recently published research in the area of assessment and present findings across an international platform in the context of the applied studio. In keeping with Kennell's (2002) work, previous research can still be grouped across several broad categories: the roles of student and teacher, their behaviors, their interactions, and evaluation. Research conducted in the 21st century can still be categorized loosely into these strands which, for the purposes of this series, have been clustered into three papers. The first paper (Parkes, 2009) examined characteristics of the applied setting and shared the work of several researchers (eg, Burwell, 2006; Colprit, 2000; Duke, 1999/2000; Duke & Simmons, 2006; Fredrickson, 2007; Gaunt, 2008; Kurkul, 2007; Laukka, 2004; Mills & Smith, 2003). The second paper, (Parkes, 2010a) focused on the use of practice in the applied studio (eg, Barry & Hallam, 2002; Byo & Cassidy, 2008; Jørgensen, 2002; Jørgensen & Lehmann, 1997; Kostka, 2002; Lehmann & Davidson, 2006; Madsen, 2004; Maynard, 2006; Miksza, 2006; Nielsen, 2001, 2008).

This third and final paper examines a body of research concerning the use of assessment and evaluation in the applied studio. Additionally, several authors (Bennett & Stanberg, 2006; Fallin & Garrison, 2005; Parncutt, 2007; Sinsabaugh, 2007) have recently highlighted a more collaborative approach to the current *modus operandi* of the applied studio. There has been some interest (Parkes, 2011, in press) in the dual roles that applied teachers undertake as performers and/or teachers which will also be explored in this paper in summary of the applied studio setting. Recommendations for the future of applied studio research will also be given. In preparing this extensive literature review for the three articles, relevant works published in English language journals were chosen to be represented, along with salient dissertations, and many databases were utilized, for example but not limited to; Academic Search Complete, Cambridge Journals Online, EBSCOhost, JSTOR, and SAGE.

Assessing musical performance

Colwell (1971) observed that applied lessons were filled with evaluation and that music schools can be seen as trade schools in regard to how they develop a set of specified skills (p. 41). Four decades later, Abeles (2010) makes a similar observation; "We listen to our students' playing, assessing their strengths and weakness, then help them to determine what additional experiences may strengthen their performance and develop them into well-rounded, independent musicians" (p. 167). Applied teachers make many assessments and evaluations, and have been doing this since the applied, or master-apprentice, lesson paradigm started in European conservatories centuries ago; clearly this is still how applied lessons operate. Faculty have given these assessments informally to students in lessons on a frequent basis. In the case of summative or formal assessments over the last decade music performance assessment research has honed a focus in a new direction, away from what has been seen by some (Brand, 1992; Jones, 1975; Madsen, 1988, 2004; Schleuter, 1997) as the secretive nature of the traditional methodologies in the applied studio.

Assessment in higher education has increased its profile since 1990 (James & Fleming, 2004-2005, p. 51) and it is an area identified by those in the measurement community as an example of unreliable measurement (Guskey, 2006). It is recognized in the 21st century that assessment provides a critical link in teaching and learning (Linn, 1994; Linn, Baker, & Dunbar, 1991; O'Donovan, Price, & Rust, 2004; Shepard, 2000; Stefani, 2004-05) and in response, researchers are exploring assessment techniques across a variety of discipline settings in higher education (eg, Brothers, 2004-2005; Brown, 2004-2005; Calvert, 2004-2005; Connor, 2004-2005; Delandshere & Petrosky, 1998; Goodnough, 2006; James & Fleming, 2004-2005; Jenkins, 2004-2005; Leinhardt, 1990; Lund & Veal, 2008; Macdonald, 2004-2005; O'Donovan, et al., 2004; Van den Berg, Admirall, & Pilot, 2006)¹. The use of assessment in the applied studio, as part of the conservatory setting, has not been a large part of the active research agendas of researchers until the past decade, although ascertaining reliable and valid measurement of musical performance was addressed in the late 1980s in the professional literature by Boyle & Radocy (1987). At that time, they reported evidence of two assessment approaches: *global* and *specific*; the former an overall impression and the latter based on specific

they reported evidence of two assessment approaches, *global* and *specific*, the former an overall impression and the latter based on specific criteria or ratings (p. 172), for evaluating music performance. They put forward the *Watkins-Farnum Performance Scale* (Watkins & Farnum, 1954) as the only the “only readily available published performance measure” (p. 174) at the time. While the Abeles (1973) applied studio performance scale may not have been published and/ or available at that time it perhaps initiated a positive increase in the number of performance scales developed for the applied studio since then (Eg, Bergee, 1988, 1989a, 1989b, 1997; Fiske, 1977; Nichols, 1991) along with the Boyle & Radocy (1987) text. Assessment tools in the applied studio today still fall into one of these two categories, that is, global or criteria-specific, and the new trajectory of investigation into more tools, rating scales, rubrics etc, shows great promise to the advancement of teaching and learning in the applied setting, especially when methodologies of the most recent decade are examined.

Tools of the 21st century

Juries and examinations

There is no question that after Abeles (1973) and Fiske (1977), the bulk of the research work carried out, in developing and testing assessment instruments, was done by Bergee (Bergee, 1988, 1989a, 1989b, 1997). Recently he and his colleagues (Bergee, 2003; Bergee & Ceconci-Roberts, 2002; Latimer, Bergee, & Cohen, 2010) have contributed further to this body of work with new studies. In 2003, Bergee proposed that “assessment of music performance in authentic contexts remains an under-investigated area of research” (Bergee, 2003, p. 137) and his study examined the inter-judge reliabilities of faculty evaluations, across brass, percussion, woodwind, voice, piano, and strings for end-of-semester jury performances from the following previously published scales (Abeles, 1973; Bergee, 1988; Jones, 1986; Nichols, 1991; Zdzinski & Barnes, 2002) creating a new one for piano. His sample of faculty participating was small (brass $n = 4$, percussion $n = 2$, woodwind $n = 5$, voice $n = 5$, piano $n = 3$, and string $n = 5$). It should be noted here that this is not pointed out as a criticism, but it is indicative of the true nature of the size of music departments and often the sample size for these types of studies. The inter-judge reliabilities, as determined with coefficients of concordance, were “good regardless of panel size” (p. 143) even between most pairwise comparisons of evaluators. Bergee reported that interjudge reliability was stable for scale total scores, subscores, and the global letter grade. Reasons given for this were that veteran faculty assisted younger less-experienced faculty in evaluating student performances. Of most interest in this study is the fact that the researcher specified to the judges participating that the students would not see the grades calculated on the criteria-specific scales. Also that the additional grade and comments, that is the traditional global and actual grading method that was also utilized by the judges, would not be compared to the rating scale grades.

This implies that faculty would not use the scales in “reality” to measure students’ musical performance, but would do so in the interest of assisting the researcher. Bergee states that in the post-evaluation discussions, judges spoke about some of the performance skills or aspects that they had not considered prior to using his scales, indicating that using the scales had made an impact on their perceptions about grading and evaluating musical performance. Parkes (2006) found faculty participation to be the largest problem with testing the use of a criteria-specific rubric designed for both brass and woodwind juries. The use of the scales resulted in Cronbach alphas of .93 each but lack of participation from applied faculty greatly diminished the potential application of the study to further ascertain faculty perceptions about grading and evaluating in the applied studio. The published outline of the full study that expounds the study’s potential, and limitations, can be found in the *Proceedings of the 2007 Symposium on Assessment in Music Education* (Parkes, 2008a)

Bergee’s work with Latimer & Cohen (Latimer, et al., 2010) investigates the reliability and perceived utility of a performance assessment rubric and while somewhat outside the scope of focus within applied settings in this paper, it is worth acknowledging as a rubric tool that has both appropriate dimensions and descriptors as well as reported improvement in pedagogical utility. The internal consistency was reported at Cronbach’s $\alpha = .88$ and item by item, the inter-correlations ranged from moderately low ($r = .46$) to moderately high ($r = .87$). Judges using the rubric were also canvassed in regards to changes they would make to the rubric and to their perceived estimation of how useful the rubric would be. Respondents reported a generally positive perception of the rubric and that suggestions centered around making the dimensions of the rubric more related or intergrated. This study illustrates a substantive connection between the work of Bergee in the applied studio into high schools, a research direction that will be addressed in more detail further in this paper.

The work of Ciorba and Smith (2009) was conducted to meet accreditation requirements of implementing specific assessment tools and showed faculty commitment to the process and school-wide use of the tool. In their study, they administered a multidimensional assessment rubric (criteria-specific) for all students performing a jury recital ($n = 359$). The results of this study showed a high degree of interjudge reliability for total scores, and less so for sub-scale scores. Most reliability coefficients were above .70, which was expected and supports the findings of Bergee (2003). However, the process by which the rubric was developed is interesting as it was not developed by one researcher, as in the work of Bergee and Parkes, but by a panel of four faculty who identified common dimensions that were shared across all areas and created descriptors outlining the expected levels of achievement. The rubrics were piloted over two semesters to refine the rubric and determine the applicability of its use in a jury, or final performance, setting. Used across all instrument and voice areas, the findings indicated that performance achievement was positively related with participants’ year in school, with a one-way multivariate analysis of variance. It would be interesting to read a follow-up study revealing what percentage of faculty continue to use the tool currently at the study site.

The above studies all have utilized measurement of the consistent aspects of performance in the studio such as technique, pitch and intonation, rhythm, musical expression or phrasing, and likewise, the recent work of Russell (2010) also reports similar categories already established in music performance structure; tone, intonation, technique, and interpretation or musical effect. It should be considered that these aspects are stable criteria in applied studio performance measurement.

Self and peer assessment

Lebler (2007) states that if the “modern conservatorium is to prosper in a rapidly changing cultural and economic landscape, it will need to ... produce graduates who are self-monitoring and self-directing in their learning” (p. 206). His argument is supported by research he carried out in Australia and while the focus is on students enrolled in a Popular Music Bachelor degree, as opposed to the traditional instrumental or

vocal, his point that “assessment practices influence the nature of student learning” (p. 207) in regard to how student evaluate themselves and others, is important to examine. His use of reflective practice, portfolio assessment, and performance assessment (Lebler, 2006) prompted this paper to take a wider review of the existing literature to establish if these techniques are used in the applied studio setting.

Parkes (2010b) conducted a very small case study for the purpose of exploring the use of self-assessment in the applied studio. One applied studio professor at a large Northern American state university agreed to use a criteria-specific performance rubric (Parkes, 2006) with her students ($n = 11$). The professor gave the rubric to the students in her studio as formative feedback in their lessons and she also instructed students to use it weekly while listening to audio recordings of their lessons. Students kept an online reflective journal in which they were to reflect about their progress, which in-turn were then micro-analyzed for themes. The following categories emerged in the findings: (1) positive perceptions about the rubric from both teacher and students, (2) an increase in student awareness and recognition of how improvements could be made after using the rubric to evaluate their own lesson performance, and (3) a clear understanding of what the professor required. Suggestions for future research included a wider use of applied faculty for the tool, as benefits appeared initially promising particularly in support of learner-centered pedagogies (Weimer, 1992).

Daniel's (2001, 2004, 2006, 2007) body of work has illustrated deeper understanding of the related aspects of this area in his research from Australia. His early work (Daniel, 2001) examines the role of the learner within self-assessment issues in the applied studio. He recognizes that “the quality of the relationship between the teacher and student will have a distinct effect on the success ... of student's absorption of the comments made by the teacher” (p. 217). Underscoring the importance of skills in self-assessment, he observes two problems with the current teaching and learning setting of the applied studio. First, “the reliance of the student on the teacher, as musical guru” and second, the reliance of students on recordings to shape and develop their skills, rather than “investigating, formulating, and reflecting on their own interpretations” (p. 218). Daniel (2001) set his study with the concert practice or group practice performance class at his institution. This class is where students perform for each other and a professor who gives graded assessment of the performance. In his experimental trial, Daniel asked students ($n = 35$) to view a video-tape of their performance, and write a 300 word “self critical reflection”. Students were asked to address several issues, namely presentation, musical accuracies, style, overall impression, progress and were also asked what their plans for future direction might be. At the end of the semester they were asked to complete a survey about the experience and additional information such as their prior experience with assessment of performance and the forms of feedback or evaluation they were used to receiving. Results reported that most students still relied on teacher comments but that they were mostly positive about the introduction of a new method of assessment. In analyzing the content of comments that students made in their 300 word reflections, Daniel found that nearly half were very critical of themselves but that 80 % felt the actual process of writing the reports increased their performance skills to some degree. Fifty-seven percent responded that it was high valuable, and a further twenty-six percent reported it was moderately valuable. These results are promising for students in Australia, where it is particularly important for higher education music education to re-evaluate their current pedagogical ethos (Daniel, 2001, p. 215).

Daniel (2004) continued this work and cites the work (p. 91) of Searby & Ewers (1997) from Kingston, England as a case-study model for peer assessment. Their work interestingly has similar categories as the above criteria-specific tools namely;

- 1) Playing skills: quality of tone; accuracy of tuning; clarity of articulation.
- (2) Observation of the musical score: appropriate tempi; attention to dynamics; clear and suitable phrasing.
- (3) Musicality and style: sensitivity to the music's period and performance practice issues; a sense of dramatic shape and pacing in the performance.
- (4) Program choice: showing variety, balance and appropriateness.
- (5) Presentation: stage awareness; rapport and communication with the audience.

(Searby & Ewers, 1997, p. 379)

Daniel (2004) investigated the effects of involving students in structured peer assessments that did not impact on final summative or weighted assessment and he examined the students' perceptions about the process. Students involved in the performance practice class were scheduled to perform twice a semester and each student was required to complete four peer evaluations each week with criteria sheets, which again included elements such as accuracy, dynamics, tone, technical control, fluency, stylistic interpretation, as well as professionalism and presentation (p. 109). These were agreed upon after several sessions of faculty and students in the first six weeks of the first semester to establish usability for the tool. These sessions were reported to have assisted in engaging students in the process of critical assessment of performance. At the end of the first semester, the majority (87%) of students ($n = 34$) reported a preference for the detailed evaluation sheet and similar results were found for the second semester ($n = 31$). Students reported that their peers were not appropriately critical. The level of criticism improved and became more accurate in the second semester. Students reported valuing the open discussion process available after the performances. Daniel suggested future refinement of the process as there was a lack of validity in the general appraisal assessments and recommended future research should examine the extent to which students' achievement in performance is impacted as a result of engaging in peer assessment (p.107).

In completing his investigation of these issues, Daniel (2007) surveyed students ($n = 40$) to summarize the nature of assessment comments provided by both faculty and peers, and to examine the differences of students between year levels and between faculty, and understand the students' perceptions in regard to validity and relevance of the assessments. He found an observable increase in the average number of comments as year level increased and that first year students gave mainly positive comments. Second and third year students gave advice and direction to their peers, whereas faculty spread the content of their comments evenly between positive, critical, and advice. Naturally, faculty gave the performers almost double the amount of comments than the students did to their peers. Daniel reports that students still value the comments of faculty more so than that of their peers, and the senior students were more accepting of critical evaluations from both peers and faculty. He notes that although the survey was intended to elicit information specifically about the assessment process, students still wanted to illustrate comments specifically related to the art of performing in their open-ended responses.

In examining differences between faculty and student evaluations, Bergee & Cecconi-Roberts (2002) explored the relationships between faculty, peer and student self evaluation of performance with a pair of experiments. In experiment one, students ($n = 29$) worked together over the semester to complete peer evaluations in groupings of similar instruments (brass or woodwind etc) each week. Peer evaluation forms from Bergee's previous study (1997) were used. At the end of the semester, students were recorded performing their instrument and were then

Bergee's previous study (1997) were used. At the end of the semester, students were recorded performing their jury performance and were then asked to evaluate themselves, as they observed their own video. Faculty also evaluated the live performance, yielding acceptable reliabilities. The peer evaluations, the students' own evaluations, and the faculty judges' evaluations were compared and there were no significant differences shown in students' ability to self-evaluate by year in school, their performance level, nor the medium. Findings revealed that peer interaction improved students' initial ability to self-evaluate but that the areas of tone and musical effect seemed to be the most difficult. In experiment 2, students ($n = 56$) self-evaluated five times and performances were recorded on a digital recorder. Students listened immediately and using revised (more detailed descriptors) evaluation categories. They also participated in group meetings again where they evaluated existing recordings of exemplary artists, along with the performances of their peers. As with experiment one, students self-evaluated their jury performance at the end of the semester, as did the judge panel. Analyses showed that over time, students' self-evaluations increased, moving toward peer evaluation scores, and away from judge scores and peer evaluation was consistently higher than judge scores.

Feedback

Duke & Simmons (2006) illustrate that goals and expectations are prominent elements in lessons given by applied faculty, in their case, world-renowned artist-teachers. The connection between these goals and expectations and the assessment points for measurement or evaluation are not specifically stated by Duke & Simmons, however the assumption of the artist-teachers in their study is that the student play in a lesson as if they are performing, to achieve "a high standard" (p. 12). This type of feedback is conveyed to the student consistently in lessons and it is reasonable to assume the student knows this expectation continues to prevail in the jury or recital setting.

The jury or recital setting is considered authentic assessment, in the literal sense of the term 'performance assessment' but there is not a great deal known about the formative effects of summative jury feedback on actual student improvement. It may be that the feedback given during lessons is what ultimately creates change in student achievement and growth.

Daniel (2006) pursued the implications of using video footage in piano teaching, with both one-to-one and group lessons. In some respects, this research speaks more to the characteristics of the applied studio, however it also illustrates the interaction between student and teacher, particularly when the issue of feedback is concerned. As such, after 150 hours of teaching were recorded and analyzed for interpretation of actions, a second round of analyses revealed classifications of behaviors that included "evaluative and diagnostic" (p. 198). Findings show that in the first set of analysis, teachers were 16 times more likely to give input than the students in the one-to-one lesson, yet only twice as likely to give input in the group lessons. In the secondary analysis data, results show that students participated minimally in diagnostics, and not at all in evaluation in the one-to-one lesson. Conversely, in the group lessons, students took a much greater role in evaluation, in fact exceeding that of the teacher. Clearly, the effects of peer interaction and participation had some causality, but this study highlights the differences in typical learning settings.

Rostvall & West (2003) also examine interaction, and, to that end also feedback in their study administered in Sweden. In instrumental applied guitar lessons (teachers $n = 4$, students $n = 21$) they studied what is still considered in Sweden to be a "hidden and secret activity that goes on behind closed doors" (p. 214). They analyzed the interactions of the group and one-to-one lesson interactions after coding hours of video footage. Findings revealed that the teachers controlled the learning settings and that they did not ask for students' perspectives at any time. Students in the one-to-one lessons were stopped or interrupted if they made spontaneous comments about the performance or music. This was not seen in the group lessons, where again peer interaction may have had an impact of the distribution of power in the room. Rostvall & West reported that when teachers ignored students perspectives they became less able to analyze the situations in which students were having problems and that students were left to self-diagnose, self-evaluate and self-correct, especially elements such as motor learning and expressive aspects of their performance. Rostvall & West additionally looked interpretively at their data and concluded that the teachers were ranked hierarchically in the setting as skilled masters, who expected only respect and obedience from students. The teachers' approach to feedback was to only impart a little of their knowledge of tradition and it was shared incrementally over time. This study illustrates that perhaps the European tradition of master-apprentice is showing marked differences, particularly in the area of feedback and evaluation, in 21st century teaching when it is considered with an international perspective.

Reliability and validity

It is clear that some researchers are establishing good reliabilities with tools used in the applied studio (Bergee, 2003; Bergee & Cecconi-Roberts, 2002; Ciorba & Smith, 2009; Parkes, 2008a) and that if used widely, these tools may make some measurable difference in the perceptions of both students and faculty about assessment in the applied studio. Asmus (1999) proposed that "greater precision in assessment will provide better information to both student and teacher, because it can help evaluation instructional strategies of the past and select appropriate strategies for the future" (p.19). He also supports teachers creating their own assessment tools assuming the validity can be assured (Asmus, 2010) and while his suggestions are made for public school music teachers, they are equally applicable to the applied studio. Most of the labor in establishing the factors present in music performance assessments has been done and reliabilities are generally strong (eg, Abeles, 1973; Bergee, 1988, 1989b; Ciorba & Smith, 2009; Fiske, 1977; H. Jones, 1986; Latimer, et al., 2010; Miksza, 2006; Nichols, 1991; Russell, 2010; Zdzinski & Barnes, 2002). As Asmus (2010) points out, the issue becomes one of validity when we consider defining the construct being measured. If it is the established elements and factors so commonly found, then are these tools appropriate to be used widely? Should they be published and be made easily and publically available?

The reactive answer is 'probably yes' and many music programs would benefit but perhaps an alternative approach could be considered because this answer raises several additional questions; Are the goals of all applied teachers the same? Are we, as performance educators, aligned across programs, schools, states, or even countries, in what we expect from our students? Is a performance rubric developed by four faculty at Milikin University (Ciorba & Smith, 2009), for example, going to be valid for use with students in a contrasting music department with another set of teachers and students, in a different college or university? It is possible then, that the answer may be "perhaps not". It might then, alternatively, be more effective to have applied faculty meet, discuss, and deliberate about the goals they have as individual teachers, and as colleagues, in order to create tools that will be firstly valid for their specific set of students, and secondly may also serve to inform teachers' practices and pedagogy. They could start with any one of the tools illustrated within this paper, and utilize it as a starting point, rather than

adopting it in totality. Tools developed within communities of teachers would most probably be reliable if the dimensions identified were central to each performance medium, the items were categorized to best represent each dimension, and they were paired with Likert-type categorical response scales (Latimer, et al., 2010). Latimer et al., also suggest that criteria-specific rating scales to be reliable due to the descriptors of specific levels of performance proficiency. This would be helpful if, for example, teachers have diverse levels of student ability and experience, such as freshman through to doctoral level performance majors. One scale or one rubric may not be appropriate. This is not to suggest that every applied faculty teacher should develop a rubric for each level of student but that by working collaboratively in departments (brass, string, voice, piano, percussion, woodwind etc) perhaps heterogeneous scales or rubrics could be created at lower division, upper division, and graduate levels. Wexler (2008) investigated goals and strategies of applied music faculty to establish current assessment practices faculty could start with this idea; establishing what they are currently doing and examining their processes for overlap and differences. Initial conversations between faculty could start where this type of baseline is established. Researchers could replicate Wexler's work and share what is occurring at their school or across several schools. It seems that a more collaborative endeavor is required between applied faculty and researchers, and as Bergee (2003) suggests, "further effort should be directed toward refining performance structures, developing valid assessments in all domains, and testing the efficacy of these assessment tools in performance contexts" (p. 148).

Implications for music education

These suggestions, if acted upon, would surely change the nature of the applied studio. There would need to be more communication between the faculty trained in educational measurement and the applied teachers. In the Northern American model this would require music education faculty, who typically have some education in research, validity, reliability, assessment, standardized testing, and statistical analyses, to participate. Parncutt (2007) advocates for a "fruitful interaction between performance teaching and performance research" (p.1) and one can widen his vision to include assessment of music performance. Visibility for music assessment has increased due to the launch of the *Symposium on Assessment in Music Education* held in Florida 2007 and 2009, and the publication of both proceedings (Brophy; 2008, 2010). The applied studio has a presence in these books, and the applied studio assessment topic has been more regularly addressed in scholarly journals over the past two decades. While this trend has added to the quality and availability of assessment tools for the applied studio, Guskey (2006) argues that "few educators [in higher education] receive any formal training in assigning marks to students' work or in grading students' performance and achievement" (p. i) and while he is not specifically speaking about applied music teachers, his point remains relevant, to some extent, to this paper. The work of Parkes (2008b) indicates that 79% of applied teachers (n = 162) who participated in her study, reported that they had no training in either pedagogical techniques, including assessment, as part of their education at all. While research has been conducted into the reliabilities, usefulness, factorial nature, and stability of these tools, the actual rate at which they are being used by applied faculty outside the research studies themselves, may be minimal. The potential effects of this may impact music education as a whole, and music performance education could be the steward of a new period of use for valid and reliable assessment techniques.

If there were to be more attention given to assessment practices as part of good instructional pedagogy, the nature of music education may also change. When one considers the most recent findings of Bazan (2010), LaCognata (2010), and even the MENC report (1998), it seems that high school teachers are still evaluating students' attitude, attendance, and participation with alarming regularity. While this is not completely inappropriate, it is a concern, particularly when students are learning important musical skills, musical knowledge, vocabulary, composition and improvising skills, as well as an understanding about music in their own, and others' cultures. It seems more appropriate to assess those elements and skills with regularity and clarity, and also communicate the importance of being able to assess them to our current music education students. Some teacher educators are addressing this issue (Bauer, 2008; Niernan, 2008; Reimer, 2009; Wuttke, 2010) and are sharing this research on an international platform, which is another positive trend. Daniel's (2006) work with higher education might also support that in high school group learning, we can improve peer evaluation processes. If the nature of assessment in the applied studio was improved, it may have a flow-on effect to the public school methods of assessment in the future.

Suggestions for future research

Some researchers are calling for more collaboration between music education and music performance (Bennett & Stanberg, 2006; Fallin & Garrison, 2005; Sinsabaugh, 2007). Sinsabaugh (2007) in particular, supports the notion that teacher-artists and educators need to communicate well with each other. Her work with partnerships between solo performers and schools gives her a unique perspective on how these communications can occur. Future research should examine best practices for "pulling together" (Fallin & Garrison, 2005) between music educators and music performers. Bennet & Standberg (2006) address the same issue from the Australian viewpoint. It would seem that the divide between music education and music performance is equally wide where "hierarchical inference ... is perpetuated in the separation of education and performance students during their university education" (p. 219). Parkes (2011, in press) addresses the issues of disjunction between performer, teacher, and educator and presents a comprehensive approach to performance pedagogy. Future research questions could address perceived cause of hierarchical perceptions, as well perceived benefits for both performer and educators to work together on the issue of music performance assessment. Parkes (2011, in press) examines this topic philosophically at the higher education level, and perhaps future researchers and authors will be encouraged to make these issues part of the dialogue occurring in K-21 music conversations with music performers and teachers.

There is an issue left untouched in this paper, and this is the area of assessing musical expression and creativity in the applied studio. How do we measure these? Is this process empirically documented or investigated? Should a student learn to approximate the expressive syntax of the master teacher or are they encouraged to create new estimations of meaning for traditional repertoire? It has been the most difficult element to assess and often has the lowest reliabilities as one dimension on currently well-tested assessment tools. There is certainly investigation into this phenomenon (Huron, 2006; Juslin & Persson, 2002, p. 229) and questions are raised as to how it is taught, so perhaps future research could establish this first and then how is it best measured?

In conclusion, this paper suggests that the nature of the applied studio is perhaps slowly changing as the current trajectory of research moves the use of valid and reliable assessments forward into more schools and programs. It is anticipated that as more research is conducted,

more assessment tools, and their use as reported in findings, will be come part of the way music performance is assessed and evaluated in the future. Practical applications may take away from this paper may be demonstrated in the following ways: (1) advocating for better communication between faculty in order to foster the goals of participating in research and removing secrecy or subjectivity in music assessment, (2) facilitating a culture of willingness to adopt or create new assessment practices in music performance schools, and (3) finally, to test and refine assessment practices to establish reliability and validity in assessment use as part of the ongoing teaching and learning cycle in the applied studio.

References

- Abeles, H. F. (1973). Development and validation of a clarinet performance scale. *Journal of Research in Music Education*, 21, 246-255.
- Abeles, H. F. (2010). Assessing music learning. In H. F. Abeles & L. A. Custodera (Eds.), *Critical Issues in Music Education: Contemporary Theory and Practice* (pp. 167-193). New York, NY: Oxford University Press.
- Asmus, E. P. (1999). Music assessment concepts. *Music Educators Journal*, 86(2), 19-24.
- Asmus, E. P. (2010). Assuring the validity of teacher-made music assessments. In T. S. Brophy (Ed.), *The Practice of Assessment in Music Education: Frameworks, Models, and Designs* (pp. 131-143). Chicago, IL: GIA Publishers.
- Barry, N. H., & Hallam, S. (2002). Practice. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 151-165). New York, NY: Oxford University Press.
- Bauer, W. I. (2008). The teacher work sample methodology: A tool to develop pre-service music teachers' understanding of assessment. In T. S. Brophy (Ed.), *Assessment in Music Education: Integrating curriculum, theory, and practice*. (pp. 363-376). Chicago, IL: GIA Publications.
- Bazan, D. E. (2010). Planning and assessment practices of high school band directors. In L. K. Thompson & M. R. Campbell (Eds.), *Issues of identity of music education* (pp. 109-125). Charlotte, NC: Information Age Publishing.
- Bennett, D., & Stanberg, A. (2006). Musicians as teachers: developing a positive view through collaborative learning partnerships. *International Journal of Music Education*, 24(3), 219-230.
- Bergee, M. J. (1988). The use of an objectively constructed rating scale for the evaluation of brass juries: a criterion related study. *Missouri Journal of Research in Music Education*, 5(5), 6-25.
- Bergee, M. J. (1989a). An investigation of the efficacy of using an objectively constructed rating scale for the evaluation of university-level single reed juries. *Missouri Journal of Research in Music Education*, 26, 74-91.
- Bergee, M. J. (1989b). An objectively constructed rating scale for euphonium and tuba music performance. *Dialogue in Instrumental Music Education*, 13, 65-86.
- Bergee, M. J. (1997). Relationships among faculty, peer, and self-evaluations of applied performances. *Journal of Research in Music Education*, 45(4), 601-612.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137-149.
- Bergee, M. J., & Cecconi-Roberts, L. (2002). Effects of small-group peer interaction on self-evaluation of music performance. *Journal of Research in Music Education*, 50(3), 256-268.
- Boyle, J. D., & Radocy, R. E. (1987). Measuring musical performance. *Measurement and evaluation of musical experiences* (pp. 171-194). New York, NY: Schirmer
- Brand, M. (1992). Voodoo and the applied music studio. *The Quarterly Journal of Music Teaching and Learning*, 3(2), 3-4.
- Brophy, T. (Ed.). (2008). *Assessment in music education: Integrating Curriculum, Theory, and Practice*. Chicago, IL: GIA Publications.
- Brophy, T. (Ed.). (2010). *The Practice of Assessment in Music Education: Frameworks, Models, and Designs*. Chicago, IL: GIA Publications
- Brothers, D. (2004-2005). A presentation: a preferred route for assessment in counsellor training. *Learning and Teaching in Higher Education*, 1(1), 90-92.
- Brown, S. (2004-2005). Assessment for learning. *Learning and Teaching in Higher Education*, 1(1), 81-89.
- Burwell, K. (2006). A comparison of the "whole" and "combination" methods of learning piano music. *Music Education Research*, 8(3), 331-347.
- Byo, J., & Cassidy, J. W. (2008). An exploratory study of time use in the practice of music majors self-report and observation analysis. *Update: Applications of Research in Music Education*, 27(1), 33-40.
- Calvert, B. (2004-2005). 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.
- Colprit, E. J. (2000). Observation and analysis of Suzuki string teaching. *Journal of Research in Music Education*, 48(3), 206-221.
- 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.
- Daniel, R. (2001). Self-assessment in performance. *British Journal of Music Education*, 18(3), 215-226.
- Daniel, R. (2004). Peer assessment in music performance: the development, trial and evaluation of a methodology for the Australian tertiary environment. *British Journal of Music Education*, 21(1), 89-110.
- Daniel, R. (2006). Exploring music instrument teaching and learning environments: video analysis as a means of elucidating process and learning outcomes. *Music Education Research*, 8(2), 161-215.
- Daniel, R. (2007). *Closing the feedback loop: an investigation and analysis of student evaluations of peer and staff assessments in music performance*. Paper presented at the The XXIXth Australian Association for Research in Music Education, Music Values, Research, and Initiatives, Melbourne, Australia July 2-4.
- Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational*

- Researcher*, 27(2), 14-24.
- Duke, R. A. (1999/2000). Measures of instructional effectiveness in music research. *Bulletin for the Council of Research in Music Education*, 143, 1-48.
- 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.
- Fallin, J. R., & Garrison, P. K. (2005). Answering NASM's challenge: Are we all pulling together? *Music Educators Journal*, 91(4), 45-49.
- Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education*, 25, 256-263.
- Fredrickson, W. E. (2007). Perceptions of college-level music performance majors teaching applied music lessons to young students. *International Journal of Music Education*, 25(1), 72-81.
- Gaunt, H. (2008). One-to-one tuition in a conservatoire: the perceptions of instrumental and vocal teachers. *Psychology of Music*, 36(2), 215-245.
- Goodnough, K. (2006). Enhancing pedagogical content knowledge through self-study: an exploration of problem-based learning. *Teaching in Higher Education*, 11(3), 301-318.
- 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 online, June 2007 http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED492005&ERICExtSearch_SearchType_0=no&accno=ED492005.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MS: The MIT Press.
- James, D., & Fleming, S. (2004-2005). 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, H. (1986). *An application of the facet-factorial approach to scale construction in the development of a rating scale for high school solo vocal performance*. Doctoral dissertation, University of Oklahoma. Dissertation Abstracts International 47, 1230A.
- Jones, W. J. (1975). Games studio teachers play. *Music Journal*, 33, 46.
- Jørgensen, H. (2000). Student learning in higher instrumental education: Who is responsible? *British Journal of Music Education*, 17(1), 67-77.
- Jørgensen, H. (2002). Instrumental performance expertise and amount of practice among instrumental students in a conservatoire. *Music Education Research*, 4(1), 105-119.
- Jørgensen, H., & Lehmann, A. C. (Eds.). (1997). *Does practice make perfect? Current theory and research on instrumental music practice*. Oslo, Norway: Norges musikkhøgskole.
- Juslin, P. N., & Persson, R. S. (2002). Emotional communication. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 219-236). New York, NY: Oxford University Press.
- Kennell, R. (2002). Systematic research in studio instruction in music. In R. Colwell & C. Richardson (Eds.), *The new handbook of research on music teaching and learning* (pp. 243-256). New York, NY: Oxford University Press.
- Kostka, M. J. (2002). Practice expectations and attitudes: a survey of college-level music teachers and students. *Journal of Research in Music Education*, 50(4), 145-154.
- Kurkul, W. W. (2007). Nonverbal communication in one-to-one music performance instruction. *Psychology of Music*, 35(2), 327-362.
- LaCognata, J. P. (2010). Student assessment in the high school band ensemble class. In T. S. Brophy (Ed.), *The Practice of Assessment in Music Education: Frameworks, Models, and Designs* (pp. 227-238). Chicago, IL: GIA Publications.
- Latimer, M. E. J., Bergee, M. J., & Cohen, M. L. (2010). Reliability and perceived pedagogical utility of a weighted music performance assessment rubric. *Journal of Research in Music Education*, 58(2), 168-183.
- Laukka, P. (2004). Instrumental music teachers' views on expressivity: a report from music conservatories. *Music Education Research*, 6(1), 45-56.
- Lebler, D. (2006). The master-less studio: an autonomous education community. *Journal of Learning Design*, 1(3), 41-50.
- Lebler, D. (2007). Student-as-master? reflections on a learning pedagogy innovation in popular music pedagogy. *International Journal of Music Education*, 25(3), 205-221.
- Lehmann, A. C., & Davidson, J. W. (2006). Taking an acquired skills perspective on music performance. In R. Colwell (Ed.), *MENC Handbook of musical cognition and development* (pp. 225-258). New York, NY: Oxford University Press.
- Leinhardt, G. (1990). Capturing craft knowledge in teaching. *Educational Researcher*, 19(2), 18-25.
- 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.
- Lund, J. L., & Veal, M. L. (2008). Measuring pupil learning- How do student teachers assess within instructional models? *Journal of Teaching in Physical Education*, 27, 487-511.
- Macdonald, A. (2004-2005). 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.
- Maynard, L. M. (2006). The role of repetition in the practice sessions of artist teachers and their students. *Bulletin of the Council for Research in Music Education*, 167, 61-72.

- MENC, Music Educators National Conference. (1998). Grading practices in music: a report of the Music Educators National Conference. *Music Educators Journal*, 84(5), 37-40.
- Miksza, P. (2006). Relationships among impulsiveness, locus of control, sex, and music practice. *Journal of Research in Music Education*, 54(4), 308-323.
- Mills, J., & Smith, J. (2003). Teachers beliefs about effective instrumental teaching in schools and higher education. *British Journal of Music Education*, 20(1), 5-27.
- Nichols, J. P. (1991). A factor analysis approach to the development of a rating scale for snare performance. *Dialogue in Instrumental Music Education*, 15, 11-31.
- Nielsen, S. (2001). Self-regulating learning strategies in instrumental music practice. *Music Education Research*, 3(2), 155-167.
- Nielsen, S. (2008). Achievement goals, learning strategies, and instrumental performance. *Music Education Research*, 10(2), 235-247.
- Nierman, G. E. (2008). The development and validation of a measurement tool for assessing students' ability to keep a steady beat. In T. S. Brophy (Ed.), *Assessment in Music Education: Integrating curriculum, theory, and practice*. (pp. 290-295). Chicago, IL: GIA Publishers.
- 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.
- Parkes, K. A. (2006). *The effect of performance grading rubrics on the college music applied studio*. Doctoral dissertation, University of Miami.
- Parkes, K. A. (2008a). 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. (2008b). The effects of grading instruction on college applied faculty perceptions about grading. *Bulletin for the Council of Research in Music Education*, 177, 79-89.
- Parkes, K. A. (2009). Recent research in applied studio instruction: Characteristics of the applied setting. *Journal of Research in Music Performance*, 1(1). Retrieved from http://www.musicalperspectives.com/Site/Recent_Research.html
- Parkes, K. A. (2010a). Recent research in applied studio instruction: practice time and strategies. *Journal of Research in Music Performance*. Retrieved from http://www.musicalperspectives.com/Site/Recent_Research-Practice_Time.html
- Parkes, K. A. (2010b). The use of criteria specific performance rubrics for student self-assessment: a case study *The Practice of Assessment in Music Education: Frameworks, Models, and Designs* (pp. 453-458). Chicago, IL: GIA Publications.
- Parkes, K. A. (2010c). Performance assessment: lessons from performers. *International Journal of Teaching and Learning in Higher Education*, 22(1), 98-106.
- Parkes, K. A. (2011, in press) College applied faculty: the disjunction of performer, teacher, and educator. *The College Music Symposium, Journal of the College Music Society*, 49-50.
- Parnutt, R. (2007). Can researchers help artists? Music performance research for music students. *Music Performance Research*, 1(1), 1-25.
- Reimer, M. U. (2009). Assessing individual performance in college band. *Research and Issues in Music Education*, 7(1). Retrieved from <http://www.stthomas.edu/rimeonline/vol7/reimer.htm>
- Rostvall, A.-L., & West, T. (2003). Analysis of interaction and learning in instrumental teaching. *Music Education Research*, 5(3), 213-226.
- Russell, B. E. (2010). The development of a guitar performance rating scale using a facet-factorial approach. *Bulletin for the Council of Research in Music Education*, 184(Spring).
- Schleuter, S. L. (1997). *A sound approach to teaching instrumentalists*. Belmont, CA: Schirmer.
- 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 and Evaluation in Higher Education*, 18(3), 221-233.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Sinsabaugh, K. (2007). *It's time for music educators and performing musicians to unite*. Paper presented at the Policies and Practices: Rethinking music teacher preparation in the 21st century, New York.
- Stefani, L. (2004-05). Assessment of student learning: promoting a scholarly approach. *Learning and Teaching in Higher Education*, 1(1), 51-66.
- 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.
- Watkins, J. B., & Farnum, S. E. (1954). *The Watkins-Farnum performance scale*. Winona, MN: Hal Leonard.
- Weimer, M. (1992). *Learner-centered teaching*. San Francisco, CA: Jossey-Bass.
- Wexler, M. (2008). *A comparative survey of goals and strategies of college music performance teachers across instrumental groups*. Unpublished Doctoral dissertation, Teachers College, Columbia University, New York.
- Wuttke, B. C. (2010). Self-assessment strategies for developing future music teachers. In T. S. Brophy (Ed.), *The practice of Assessment in Music Education: Frameworks, Models, and Designs* (pp. 391-394). Chicago, IL: GIA Publisher.
- Zdinski, S. F., & Barnes, G. V. (2002). Development and validation of a string performance rating scale. *Journal of Research in Music Education*, 50(3), 245-255.

