

The Evaluation of Music Faculty in Higher Education: Current Practices

Kelly A. Parkes

Teachers College Columbia University

The purpose of this research was to ascertain the methods used to evaluate music faculty and whether achievement measures, or student progress, impact the evaluations made about teacher effectiveness for music faculty in the higher education context. The author surveyed Chairs of Departments or Directors of Schools of Music (n = 412) listed as degree-granting (Baccalaureate, Masters, and Doctorate) in music performance on the National Association of Schools of Music's current membership directory in the United States. Administrators (n = 142) responded to an emailed link to an online survey where they were asked to give information regarding their programs, their faculty, and their processes for evaluating teachers' effectiveness, yielding a response rate of 34%. Methods of faculty evaluations and the ways in which they were used were examined. Respondents shared exemplars of the instruments used to evaluate faculty. Results from this study suggest that the methods for evaluating faculty include students' perceptions of instruction, peer evaluations of teaching, self-assessments of teaching and measures of student progress as the current practices being employed. Suggestions for the field include further investigation as to what administrators might agree upon as to appropriate measures of student progress, achievement or growth.

Evaluation, in general, in higher education has become nationally important and is recognized as one of the ways in which teacher effectiveness can be improved. The overarching philosophy is that if teaching effectiveness or quality can be measured, then teachers needing assistance can be identified and remediated, while highly effective or high quality teachers can be rewarded within the promotion, tenure and merit systems. Many institutions of higher education across the USA include an Office of Assessment on campus, and other similar centers or initiatives, not only to assist in accreditation procedures and to provide measures for accountability, but also to garner data regarding the teaching quality of professors. In some cases this identifies certain teachers as having more or less impact on student learning. Institutes for Teaching and Learning are also part of many higher education institutions, and they are responsible for providing professional development for faculty, as well as measuring levels of student perceptions about the instruction they have received. Institutions differ in regards to which measures are used in evaluating faculty teaching performance, but traditionally, student perceptions of instruction surveys have been used along with peer evaluations of teaching. To determine whether practices have changed as higher education moves into the 21st century, this paper examines how performing arts faculty are currently evaluated in partial replication of a previous study (Hipp, 1979). Hipp's extensive dissertation focused on the evaluation of music faculty in regards to several factors, such as promotion, tenure, retention, teaching effectiveness, faculty development, merit increases, teaching assignments and committee assignments. The current study focuses only on evaluation of teaching effectiveness.

Literature Review

The recently published American Educational Research Association report (AERA, 2013) recommends, based on the work of a task force, that to evaluate teaching the focus should be on student learning outcomes (p. 1). More specifically, the report provides recommendations that systems designed to evaluate faculty would assist institutions define 'teaching quality' based on student learning outcomes. It is recommended that faculty members be assisted to improve their teaching by identifying where they need professional development, and that evaluators determine a faculty member's relative strengths and weaknesses (p. 3). In the field of the arts, the National Association for Schools of Music also make available a document (NASM, 2009) designed to assist institutions, programs, and individuals making local assessments regarding arts faculty evaluation and reward systems in higher education. This document, based on the work of an interdisciplinary task force, encourages the user to determine the primary indicators of merit with respect to teaching, service, creative work and research (p. 10) and to understand which perspectives might be critical in determining merit. The document also suggests that arts units determine which opportunities are available to faculty in terms of support, time and peer review (p.12). Given that institutions of higher education determine their own expectations for promotion and tenure, there will be observable differences. Examining the Education document and the Arts document illustrates this clearly, and, naturally, different disciplines should have different expectations. From a thorough review of the literature, however, it has become evident that little is known about how arts, and specifically music, faculty are currently evaluated for their teaching. Given that the arts often have very different teaching and learning

settings, it seems that further investigation may be warranted.

For over 25 years, there have been at least two standard procedures used in evaluating faculty teaching in general (Kulik & McKeachie, 1975; Seldin, 1999). The first procedure relies on observations of the quality of teaching. Observations of teaching behaviors have been made by students, colleagues, supervisors or, in some cases, the teacher himself (Kulik & McKeachie, 1975, p. 210). The second type of procedure for evaluating faculty involves measures of students' performance, and it is here the music performance and pedagogy research literature falls short in regards to evaluation in the music performance courses in higher education. It is hard to find significantly real differences among grades awarded to performing arts music students; therefore, it becomes difficult to relate small differences to any characteristic of the teacher. Researchers (for example, Abeles, 1975; Duke & Simmons, 2006; L'Hommedieu, 1992; Parkes & Wexler, 2012) have established which characteristics and behaviors are effective in applied music teachers (teaching in the one-to-one studio setting), but we do not know if these characteristics and behaviors impacted cumulative learning in students over time or what other characteristics are expected in large-group classrooms such as music history or theory. We do not know what is measured in peer observations of teaching nor whether these observed behaviors are valued by administrators when they evaluate the teaching effectiveness of applied music performance teachers (Hipp, 1983).

The American Association of University Professors (AAUP, 2014, p. 201) makes a distinction between student perceptions and student learning as two different types of data when assessing the effectiveness of instruction. Older studies (e.g., Rodin & Rodin, 1972) in the wider education literature suggest that the positive correlations between students' ratings of a teacher and the achievement (grades) of that teacher's students are a pitfall of using student observations of instruction as the only measure of a teacher's effectiveness. Corroborating or triangulating several measures of teaching effectiveness (e.g., with formal observations, peer assessment, self assessment, and student evaluations) at multiple points in time may give a more comprehensive picture as to the strengths and weaknesses of a teacher's effectiveness or competency. Publications such as Berk's *Thirteen Strategies to Measure College Teaching* (2006) might be a useful example of such multiple measures; the bulk of this publication is designed to assist faculty, administrators and clinicians in developing rating scales across a variety of evidence sources such as student ratings, peer ratings, external expert ratings, self-rating, videos, student interviews, exit and alumni ratings, employer

ratings, administrator ratings, teaching scholarships, teaching awards, learning outcome measures and teaching portfolios. He suggests using national professional standards (Standards for Educational and Psychological Testing (AERA, APA, & NCME Joint Committee on Standards, 1999) for how teaching effectiveness or performance should be measured (p.12) to move beyond simply using student ratings. Other models of faculty evaluation (e.g., Arreola, 2000; Braskamp & Ory, 1994; Centra, 1993; Keig & Waggoner, 1994; Romberg, 1985) also include multiple sources of evidence with more weight given to student evaluations and peer evaluations. Berk (2006) stresses the importance of field testing and item analyses when using rating scales to measure teaching effectiveness. His main concern is with item descriptive statistics (p.148), interitem and item-scale correlations (p.152) and factor analysis (p.155); in general, he advocates the importance of collecting evidence of validity and reliability (p. 161-182). Clearly internal consistency is important when using scales or instruments to measure teaching effectiveness, and a lack of internal consistency affects the usefulness of any instrument.

The current study explores how some performing arts faculty are currently being evaluated and answers the following research questions: 1) How are music faculty in higher education in the United States currently being evaluated for their teaching? 2) Have the reliabilities and validities of the instruments used to measure teaching effectiveness been examined? 3) Are measures of student learning outcomes or progress part of the process? If so, what are those measures? 4) Have the reliabilities and validities of the measures of student outcomes or progress been examined?

Method

The survey was developed by the author and uses items (with permission, personal communication, December, 2012) from Hipp's (1979) dissertation focused on the evaluation of music faculty. The author used only the teaching effectiveness items from Hipp's study to determine how music faculty are being evaluated today. The instrument used was a small subsection of Hipp's original survey, and the items can be seen in Appendix A. An invitation was sent to 412 directors to take an online survey, and 142 completed the survey, giving a response rate of 34% in total. The survey invitation and two subsequent invitations were sent to the Chairs of Departments, or Directors of Schools of Music listed as degree-granting (Baccalaureate, Masters, and Doctorate) in music performance on the National Association of Schools of Music's current membership directory in the United States. Administrators responded to an emailed link to the online survey where they were asked to give

information regarding their programs, their faculty and their processes for evaluating teachers' effectiveness. The survey responses were descriptively and qualitatively analyzed and are represented in the following section.

Results

The respondents firstly described their program, institution, school and position, and they answered questions regarding the specifics of how music faculty are evaluated. The descriptive data illustrating their official capacity or position is shown in Table 1. Administrators represented publicly funded institutions ($n = 94$, 66%) and privately funded institutions ($n = 48$, 34%), and their institutions offer the following highest degrees: undergraduate, ($n = 47$, 33%) masters ($n = 77$, 54%), and doctorate ($n = 18$, 13%). Their music major and institution enrolment sizes can be seen in Table 2.

Most units 97% ($n = 138$) had written policies pertaining to the evaluation of music faculty and 3% ($n = 4$) did not. Of these institutions with policies, 59% ($n = 81$) were institution-wide and 41% ($n = 57$) were policies developed specifically for the music unit. The same 41% ($n = 57$) reported that music faculty participated in the formulation of these policies when they were developed specifically for the music unit. Most respondents 82% ($n = 117$); however, reported that their units also used institution-wide evaluation instruments for faculty evaluation. Of these 117 respondents, only 37 (32%) reported that this instrument was developed by an institutionally provided center for the support of teaching in higher education.

Faculty Evaluation Instruments

Fifty six percent ($n = 79$) of respondents reported that their units utilized instruments that had been specifically designed for the evaluation of music faculty, and of these, 49% reported using a single form for all types of music instruction, 54% reported using a form specifically for applied studio teachers, 42% reported using a form specifically for the evaluation of ensemble directors, 6% reported using a form specifically for the evaluation of composition / theory teachers, and 51% reported using a form for the general evaluation of classroom teachers. Administrators were asked to further describe the forms, and open-ended responses were grouped into the following categories: additional miscellaneous details, administration timelines, and developmental processes as to how forms were developed. Examples are illustrated in Table 3.

When asked if their faculty evaluation form had been examined for internal consistencies, 18 (24%)

respondents reported in the affirmative, and 58 (76%) reported that the form used for music faculty had not been examined. Of the 24% that reported examination of internal consistencies, the following processes were described: general consultation with faculty, use of standard deviation calculation, faculty vote, cross-checking, campus consultants, review by evaluation committee, data tracked by Director of Institutional Assessment, and internal SACS accreditation committees.

Student Evaluations of Faculty

Most administrators (98%, $n = 138$) reported the use of student evaluations of faculty. In terms of the specific types of student input used, the respondents reported the following: course/instructor surveys and questionnaires (100% of participants), personal statements from students (65% of participants), student reference letters (12% of participants) and other types such as student comments, interviews with students, personal statements from students and observations of committee mentors (3% of participants). Sixty-eight percent of administrators reported no examination for internal consistencies. Thirty-two percent ($n=44$) of administrators reported that their student surveys or questionnaires had been examined for reliability or validity (internal consistencies). The processes for examining internal consistencies were similar to the faculty evaluation forms and included processes such as cross-checking, review committees, institutional evaluations, internal assessments by a Director of Assessment, evaluation committees, comparisons to national data banks, evaluation by a research center on campus, Office of Institutional Research examinations, faculty senate examinations and institutional verifications. One respondent made a point of explaining that their student rating form, while internally consistent and developed by a leading psychologist, was not a good fit for their music students (Participant 20.9c). Nearly all administrators (91%) reported that teachers cannot opt out of participating in student ratings / evaluations.

Peer Evaluation

Most (81%) respondents reported that peer evaluation was utilized in their music units. The formats of peer evaluation included (but were not limited to only one) formats such a narrative report based on an observation (85%), evaluation forms (40%), reference letters (39%), personal statements from peers (38%) and questionnaires (5%). Evaluation of teaching conducted by professionals outside the institution was conducted by 41% of the music units. Of those, similar formats of input were sought; reference letters (81%), personal statements (36%), evaluation forms (14%) and questionnaires (3%) from peers outside the institution.

Table 1
Official capacity

Title	n	Percent (%)
Head of Department or School	33	23
Chair of Department or School	78	55
Associate Dean	3	2
Dean	20	14
Other*	8	6

Note: *Other responses were Director of School ($n=4$), Chair of Music Division ($n=3$), Coordinator of Department ($n=1$)

Table 2
Enrollment Size Ranges

Major Enrollment	Number	Percent (%)
Under 50	9	6
51-100	31	22
101-200	43	30
201-400	42	30
401-600	11	8
601-700	5	4
Above 700	1	1
Institution Enrollment		
Under 2500	23	16
2501-7499	27	19
7500-10,000	14	10
10,001-20,000	41	29
20,001-30,000	24	17
30,001-40,000	10	7
40,001-50,000	3	2

Self-evaluation and Evaluation by Alumni

High levels (75%) of self-evaluation were reported to be used, and examples of the types of formats shared were self-reflective narratives regarding growth over time (93%), student scores or measures of student achievement (59%), quality teaching in videos (49%), and other types of evidence (25%) such as “supporting documentation, examples of syllabi, assignments, student work, course documents, teaching portfolios, student letters, examples of student achievements, examples of students meeting learning objectives, students’ placements in graduate programs and other student awards.” Administrators (11%) reported that they used evaluations from alumni in evaluating faculty, but when used, questionnaires (44%) were the most often solicited, along with and reference letters (38%) and personal statements (19%).

Student Progress

Student progress measures were used specifically as part of faculty evaluations at 52% of the music units

in this study. Student progress measures were reported to consist of (but were not limited to) standardized tests (10%), pre-post-tests (11%), departmental examinations (42%), grade distributions (22%), informal (78%) and other types (30%) such as “performance observations, tracking of graduates, jury exams, graduation rates, performance awards, competition and job placements, student performance in ensembles and other reports by unit Chairs or Heads”. The majority of administrators (93%) reported that the student progress measures had not been examined for internal consistencies. The 7% of administrators who did examine for internal consistencies reported processes such as continual review by peers, faculty committee review, college department review and statistical procedures.

Administrators were asked about the progress or achievements of former students being included in faculty evaluation. Sixteen percent of administrators reported using them. In particular, administrators reported informal assessments of former students (71%), along with job placements records (43%), questionnaire (10%) and on-line surveys (5%). Administrators explained that, “reports of student achievements are

Table 3
Categories of additional descriptions

Descriptor examples	Frequency
Details	11
Separate evaluation forms	
No specific form	
Narrative evaluation	
Scantron-type evaluations	
The form is simply a comment sheet	
Peer evaluation form	
Course evaluation form	
Music-unit specific form	
The form measures three categories: teaching, professional activities, and service	
Administration	28
Administered online	
Administered at the end of semester	
All faculty are evaluated each semester	
Used by the Department Chair at the end of semester	
Questionnaire can be used for peer evaluation as well	
Administered in Fall and Spring semesters.	
Administered annually	
Development	26
Faculty developed the form	
Developed by faculty	
Developed by music faculty over a long period of time	
Started with templates from MTNA ... areas adapted them	
Developed by music department chair	
Developed as part of Retention, P & T criteria and vetted at the Institutional level	
Developed a long time ago	
Developed by the college, applied by the music unit	
Developed by the full-time faculty	
Questionnaire developed by School of Music and Dance	

supplied by the faculty member.” They also reported encouraging faculty to list the accomplishments, career success and current positions of former students.

Additional Observations of Teaching

Thirty five percent of administrators reported using additional observations of teaching and of those, they described the observations occurring by Deans, or Directors or other administrators at regular intervals, but especially in the case where faculty were coming up for tenure decisions. Other observations were reportedly made by the University Teaching Excellence Center or equivalent. Administrators were asked if they evaluated their applied studio teachers differently to their other faculty and thirty-three percent said they did. The results of how applied studio teachers are evaluated differently are reported elsewhere (Author, in press),

but more than half the administrators (67%) evaluated all their music faculty with the same criteria.

Importance and Sources of Evaluations

Administrators were asked to rank the various methods of evaluation of faculty, e.g., evaluations by students, by colleagues, by alumni, by outside professionals, self-evaluations, students' progress, former students' progress or observations of teaching (on a scale of 1-5, 1 having no importance and 5 having extreme importance). In this study, administrators ranked the evaluations by students and by colleagues (peers) the highest. Table 4 illustrates all the rankings of the ways in which applied faculty may be evaluated.

Administrators were also asked to rank how important (with the same scale) the ways in which the

Table 4
Rankings

Methods of Evaluation	Mean	SD
Evaluations by colleagues	4.13	0.99
Evaluation by students	4.08	0.74
Self evaluation	3.73	1.19
Other observations of teaching	3.27	1.21
Evaluation of students' progress	3.11	1.21
Evaluation by outside professional sources	2.67	1.38
Evaluation of former students' progress	2.06	1.09
Evaluation by alumni	1.88	0.88

Note: $n = 142$ for this question.

results of faculty evaluations were used. Table 5 illustrates their rankings.

Administrators were additionally asked to choose the single most important use of faculty evaluations, and their answers show that making decisions for tenure (37%) and improving teaching effectiveness (31%) were the most important.

Teaching Evaluation Tools

Several ($n = 13$) administrators chose to share their actual forms, tools and rating scales as part of this research study. An analysis of the documents revealed some commonalities and similarities. The most common tool shared was an observation form. This type of tool listed procedures and behaviors that were expected before and during teaching, such as planning instruction and assessments, setting objectives and engaging in good teaching methods. Teaching methods varied widely. Elements such as being organized, making students aware of the goals of the course, engaging students in meaningful participation, communicating clearly, demonstrating enthusiasm, having command of the subject matter/ course material, using class time effectively, responding appropriately to student questions, encouraging critical thinking, providing clear explanations, being available outside of class meetings, dealing with topics in an interesting manner, having a degree of rapport with students and providing student with feedback after assessing achievement appropriately were all included. The format of some documents varied with several allowing space for comments in the above areas while others were in checklist form on which the observer checks off observed behaviors. Some forms required description only of class/lesson activities and then allowed space for a narrative describing the teaching effectiveness. It is important to note here that this analysis was only conducted on the 13 tools that were shared. These tools most likely do not represent all the types of teaching evaluation tools of the sample of all respondents.

Administrators also shared peer evaluation and self-evaluation forms, which asked faculty to rate their perceived levels of effectiveness in helping students learn. One peer evaluation form was designed in Likert-type scale to which the peer-observer could respond to prompts such as “the instructor was well organized and prepared” and “the instructor maintained a good balance between technical and musical concerns” with responses from Strongly Disagree to Strongly Agree.

Discussion

Generally, the findings of this study illustrate that music faculty, as represented by the respondents to this survey, are evaluated primarily with student evaluations of teaching / instruction (98%), with peer evaluations of teaching (81%), with self-evaluations of teaching (75%) and measures of student progress (52%). This is not dissimilar from the results of Hipp (1979), and a comparative table (Table 6) illustrates the differences between the current study and Hipp’s data with respect to the types of evaluations used. It seems that peer and self-evaluations of teaching have increased in use, and perhaps are valued more today than in 1979 in regards to determining faculty teaching effectiveness.

The first research question sought to determine how music faculty in higher education in the United States are currently being evaluated for their teaching. From the responses from these administrators, it seems that over a half (56%) use evaluation tools designed by the music unit faculty specifically for the music faculty. These measured their teaching along with institutionally required teaching evaluation measures such as student evaluations. It is encouraging to report in the current study that 97% of units have written policies and procedures for faculty evaluation. The Hipp (1979) study reported that only 76% of music units had policies and procedures in place for evaluating faculty, and that 58% of these were developed specifically for the music unit. Figure 1 illustrates the wide variety in types of tools/instruments being used, as well as variety

Table 5
Rankings of how results are used

Evaluations are used...	Mean	SD
To make decisions regarding promotion	4.58	0.73
To make decisions regarding tenure	4.56	0.89
To make decisions regarding retention	4.32	0.89
To improve teaching effectiveness	4.29	0.76
To encourage faculty development	4.03	0.80
To formulate individual faculty goals	3.93	0.85
To make decisions regarding merit increases in salary	3.62	1.53
To make decisions regarding teaching assignments	3.35	1.09
To make decisions regarding committee assignments	2.06	1.02
To make decisions regarding class scheduling	1.98	1.06

Note: $n = 142$ for this question

Table 6
Comparison of Current and Hipp's Data

Types of Evaluations	Current (N=142)	Hipp (N=364)
Student evaluations of teaching / instruction	98%	91%
Peer evaluations of teaching	81%	57%
Self-evaluations	75%	38%
Student progress	52%	55%

in the processes used to create the tools/instruments. Twenty-four percent of administrators reported that these tools/instruments used for music faculty evaluation in general had been examined for internal consistencies, and while the exact reasons for this are not known, it may be because there are several different types of instruments being used and some do not lend themselves easily to an internal consistency analysis.

The second research question of the current study inquired as to the reliabilities and validities of the instruments specifically used to measure teaching effectiveness, such as student evaluations. Thirty-two percent of administrators reported their student evaluation instruments had been examined for internal consistencies, and the remaining 68% reported not examining for internal consistencies. Again, it is not clear as to the reasons for this; it could be the case that administrators who reported not examining for internal consistencies might be unaware about the processes, or that they had been performed by another office on campus, or that there were other reasons not explored by the current study. Obviously this finding warrants further investigation. Respondents reported that instruments had not been examined in the case of peer and self-evaluation, nor alumni evaluation. Peer evaluation processes such as narrative reports and descriptive writings are difficult to examine in terms of consistency; however, in the case of the one Likert-type

peer observation scale that was shared by a respondent, findings from this study support a move for administrators to start considering examination of internal consistencies where appropriate for items such as Likert-type or ranking scales.

Research question three was concerned with whether measures of student learning outcomes part of the evaluation process for faculty. Student progress measures are indeed being used by more than half (52%) of the respondents in this study. A variety of measures were used including standardized tests (10%), pre-post-tests (11%), departmental examinations (42%), grade distributions (22%), informal measures (78%) and other types. The other types such as performance observations, tracking of graduates, jury exams, graduation rates, performance awards, competition and job placements, student performance in ensembles and other reports by unit Chairs or Heads which might provide deeper insight into whether a student has reached their full potential with a teacher. This study also reveals that former student progress, as reported by faculty, is also used by 16% of units. Hipp (1979) reported that 27% of units were using former student progress measures and that 55% of units used student learning outcomes; however, they were described and utilized in different ways. Hipp (1979) reported only four percent of units used standardized tests, 2% used the pre-post test method, 13% used departmental

examinations, 43% used jury examinations, 11% used grade distribution data, and 45% made informal assessments about a student's progress. It seems that music units today are still making informal assessments about students' learning outcomes and are looking less often to the achievements of former students as an indicator of teacher effectiveness.

Research question four was regarding the reliabilities and validities of the measures of student learning or progress. While only half the respondents in this survey use measures of student learning outcomes or progress in their faculty evaluations, it is clear that this type of data is not examined for internal consistencies. The data that illustrates music student success is perhaps not to be found in a test score of some kind, but rather in a series of data points that show a trend. For example, if a student gets good grades, plays well in ensembles, graduates on time, and garners a position in a good graduate program or a job placement, the data points support a conclusion that this student was successful.

Given the breadth of goals music units have for their students upon graduation, such as skill acquisition, development of a performer identity, nuanced musicianship, a high level of performance, and graduate school placement or job placement, the question raised is, are the markers being used as evidence of student learning outcomes and/or progress enough or appropriate? Also, are the data points current administrators are using actually indicative of good teaching? This leads to a further research question that is raised by the inclusion of mixed student progress data points in the teaching evaluation of music faculty, and that is, which of these data points are attributable to the teacher? Finally, how do administrators find a fair, defensible strategy for combining these multiple sources of information to make evaluations about a teacher's effectiveness? This question should be examined in the near future with empirical research.

Berk (2006, p.13) suggests that these multiple sources can "serve to broaden and deepen the evidence based used to evaluate courses and to assess the quality of teaching," however, he underscores the importance of a unified conceptualization of teaching effectiveness for higher education in general. The AERA report also recommended that in evaluating teaching the focus should be on student learning outcomes (AERA, 2013, p. 1). Music units are in a unique position where the student learning outcome goals can vary from unit to unit, depending on the degrees offered. Music unit administrators meet regularly at the National Association for Schools of Music meetings and as such, it is recommended that an open discussion be tabled to outline appropriate goals and data points. Standardization is not the goal; instead, there should be the identification of a series of appropriate goals and

data points for small units, for large units, for public or privately funded institutions, for university departments and for music schools. Institutions who pursue this could make use of the Degree Qualifications Profile (Lumina Foundation, 2014) and explore the differences between specialized knowledge, broader integrated knowledge, intellectual skills, applied learning, civic learning and institution-specific areas to align the expected learning outcomes for students.

An alignment of student learning outcomes and degree expectations could naturally extend to an outlining of what is expected of teachers. This notion of a "unified conceptualization of teacher effectiveness" (Berk, 2006, p. 13) is a worthwhile suggestion and should be useful for other disciplines as well. Berk suggests additionally that unit administrators develop their own rating scales for evaluating teaching and courses, and he illustrates in detail the techniques for doing so, as well as for undertaking the necessary reliability and validity testing. The answers to research question two of this study illustrate that; in particular, many music units (68%) may not know the reliability and validities of the student evaluation instruments they are using to measure their faculty teaching effectiveness. Given that 98% of music units in this study use student evaluations as one of the primary measures of teaching effectiveness, this seems concerning and may also be of concern for other disciplines that find themselves in a similar position. Music faculty in this study have clearly been involved in developing, writing and examining the instruments used to evaluate teaching effectiveness, but perhaps they need more assistance from their institutional centers of assessment to determine whether these instruments are evaluating the constructs intended and whether they are doing so in a consistent, valid and reliable manner. This is especially important when the use of faculty evaluation in teaching is for high-stakes decisions such as promotion and tenure.

References

- Abeles, H. (1975). Student perceptions of characteristics of effective applied music instructors. *Journal of Research in Music Education*, 52(3), 248-263. doi: 10.2307/3345287
- American Association of University Professors. (2014). *Statement of teaching evaluation*. Retrieved from <http://www.aaup.org/report/statement-on-evaluation>
- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. (2nd ed.) Bolton, MA: Anker.
- AERA (2013). *Rethinking Faculty Evaluation: AERA report and recommendations on evaluating education research*,

- scholarship, and teaching in postsecondary education*. Retrieved from http://www.aera.net/Portals/38/docs/Education_Research_and_Research_Policy/RethinkingFacultyEval_R4.pdf
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) Joint Committee on Standards. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Berk, R. A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, VA: Stylus Publishing.
- Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work*. San Francisco, CA: Jossey-Bass.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.
- 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.
- Hipp, W. J. (1979). *Practices in the evaluation of music faculty in higher education*. Unpublished doctoral dissertation, University of Texas, Austin.
- Hipp, W. J. (1983). *Evaluating music faculty*. Princeton, NJ: Prestige Publications.
- Kieg, L. W., & Waggoner, M. D. (1994). *Collaborative peer review: The role of faculty in improving college teaching*. (ASHE/ERIC Higher Education Report, No. 2) Washington, DC: Association for the Study of Higher Education.
- Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. *Review of Research in Education, 3*, 210-240.
- L'Hommedieu, R. L. (1992). *The management of selected educational process variables by master studio teachers in music performance*. Doctoral dissertation, Northwestern University, Chicago, IL.
- Lumina Foundation (2014). The degree qualifications profile. Retrieved from http://www.luminafoundation.org/publications/The_Degree_Qualifications_Profile.pdf
- NASM (2009). Local assessment of evaluation and rewards systems for arts faculties in higher education. Retrieved from http://nasm.arts-accredit.org/site/docs/PUBLICATIONS/Local_Assessment_of_Eval_and_Reward_Systems-Feb2009.pdf
- Parkes, K. A., & Wexler, M. (2012). The nature of applied music teaching experience: Common elements observed in the lessons of three applied teachers. *Bulletin for the Council of Research in Music Education, 193*, 45-62. doi: 10.5406/bulcoursemsedu.193.0045
- Romberg, E. (1995). Description of peer evaluation within a comprehensive evaluation program in a dental school. *Instructional Evaluation, 8*(1), 10-16.
- Seldin, P. (1999). Current practices-good and bad-nationally. In P. Selding & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions*. (p. 1-24). Bolton, MA: Anker.

KELLY A. PARKES, Ph.D. is currently an Associate Professor in the Music and Music Education program at Teachers College Columbia University, however this research was conducted while she was an Associate Professor at Virginia Tech. Her primary research interests are in music and music education assessment, measuring aspects within the applied studio, teaching readiness, professional dispositions, and reflective practices in pre-service teachers, in addition to teaching effectiveness and teaching quality.

Appendix A
Instrument Items, replicated from Hipp (1979)

1. Please indicate your official capacity or title
2. Please indicate whether your institution is state or privately supported
3. Please indicate the highest degree offered by your music unit
4. Please indicate the enrollment range within which your institution falls
5. Please indicate the enrollment range within which your institution falls
6. Are there written policies pertaining to the evaluation of faculty in your institution? (Y/N)
 - a. If so, are these policies and procedures institution-wide, or were they developed specifically for the music unit?
 - b. If so, are these policies and procedures institution-wide, or were they developed specifically for the music unit?
7. Does your music unit use faculty evaluation instruments that are utilized institution-wide? (Y/N)
 - a. If so, is this instrument developed by an institutionally provided center/ institute for the support of teaching in higher education? (Y/N)
8. Does your music unit utilize evaluation instruments that have been specifically designed to the evaluation of music faculty? (Y/N)
 - a. If so, please indicate the type or types of instruments in use. Check more than one item, if applicable. (Single form for all types of instruction, a form for the evaluation of applied studio music teachers, a form for the evaluation of ensemble directors, a form for the evaluation of composition/theory teachers, a form for the evaluation of classroom teachers)
 - b. Has this, or have these instruments, been examined for their internal consistencies, such as reliability and validity? (Y/N)
 - c. If so, please explain the process by which internal consistencies were determined.
9. Are student ratings / evaluation of faculty engaged in within your music unit? (Y/N)
 - a. If so, please indicate the types or types of student input used. Check more than one item, if applicable. (Course/instructor surveys, personal statements from students, student reference letters)
 - b. Has the survey or questionnaire instrument been examined for their internal consistencies, such as reliability and validity? (Y/N)
 - c. If so, please explain the process by which internal consistencies were determined.
10. Does the instructor have the option of participating or not participating in student rating (evaluations) of his/her teaching? (Y/N)
11. Is evaluation by colleagues (peer evaluation) engaged in within your music unit? (Y/N)
 - a. If so, please indicate they type or types of input that are used for peer evaluation. Check more than one item, if applicable. (An evaluation form, a narrative report based on an observation, questionnaires, reference letters, personal statements)
12. Is evaluation by professionals **outside** the institution a part of the faculty evaluation process in your music unit? (Y/N)
 - a. If so, please indicate the type or types of input used by the professional outside the institution. Check more than one item, if applicable. (An evaluation form, questionnaires, reference letters, personal statements)
13. Is the faculty member being evaluated required to provide a self-evaluation of his / her own teaching? (Y/N)
 - a. If so, please indicate the type, or types, of input a faculty member might provide in a self evaluation. Check more than one item, if applicable (Examples of quality teaching – e.g. videos, self reflective narrative regarding growth over time, student scores/ measures of achievement, other – please explain)
14. Is evaluation by alumni a part of the faculty evaluation process in your music unit? (Y/N)
 - a. If so, please indicate the type or types of input provided by alumni. Check more than one item, if applicable. (An evaluation form, questionnaires, reference letters, personal statements)
15. Is an assessment of the progress of a faculty member's students a part of the faculty evaluation process in your music unit? (Y/N)

- a. If so, please indicate the form that these student progress assessments take. Check more than one if applicable. (Standardized tests, pre-test post-test, departmental examinations, grade distributions, informal, other – please describe)
16. Has this, or have these assessments, been examined for internal consistencies, such as reliability and validity? (Y/N)
 - a. If so, please describe the process by which internal consistencies were determined.
17. Is an assessment of the progress or achievement of a faculty member's **former** students a part of the evaluation process in your music unit? (Y/N)
 - a. If so, please indicate how these assessments of former students are made. Check more than one item, if applicable. (Questionnaires, online surveys, job placements records, informally, other – please describe)
18. Are any other observations made of a faculty member's teaching, by individuals such as Administrators or Centers for Learning and Teaching in Higher Education? (Y/N)
19. Please indicate the relative importance (1 is not important, 5 is extremely important) of the following sources and methods of evaluative input regarding teaching effectiveness utilized in your music unit for decisions affecting the awarding of promotions in rank, tenure, and merit increases in salary for **all music faculty that teach groups of students** (e.g., classroom teachers, composition /theory teachers, ensemble directors) (Evaluation by students, evaluations by colleagues, evaluation by alumni, evaluation by outside professional sources, self evaluation, evaluation of students' progress, evaluation of former students' progress, other observations of teaching)
20. Please indicate the relative importance (1 is not important, 5 is extremely important) of the following sources and methods of evaluative input, regarding teaching effectiveness utilized in your music unit, for decisions affecting the awarding of promotions in rank, tenure, and merit increases in salary for **applied music performance faculty** (Evaluation by students, evaluations by colleagues, evaluation by alumni, evaluation by outside professional sources, self evaluation, evaluation of students' progress, evaluation of former students' progress, other observations of teaching)
21. Please indicate the relative importance (1 is not important, 5 is extremely important) of the ways in which the results of faculty evaluation are currently used in your music unit. (To encourage faculty development, to improve teaching effectiveness, to formulate individual faculty goals, to make decisions regarding tenure, to make decisions regarding promotion, to make decisions regarding merit increases in salary, to make decisions regarding teaching assignments, to make decisions regarding committee assignments, to make decisions regarding class scheduling.)
22. Would you be willing to share the instrument or instruments your unit uses to evaluate music faculty teaching effectiveness? (Y/N)