A facet-factorial approach towards the development and validation of a jazz rhythm section performance rating scale

Brian C. Wesolowski
The University of Georgia, USA

Abstract

The purpose of this study was to develop a valid and reliable rating scale to assess jazz rhythm sections in the context of jazz big band performance. The research questions that guided this study included: (a) what central factors contribute to the assessment of a jazz rhythm section? (b) what items should be used to describe and assess a jazz rhythm section performance? (c) how should the items be categorized? (d) what differences among jazz rhythm sections exist at three performance achievement levels? and (e) what criteria best predict group membership into three performance achievement levels? Items were gathered from research and literature related to the assessment, teaching, and general discussions related to the jazz rhythm section. Twenty-nine item statements were paired with a four-point Likert scale. One hundred and twenty-two responses were gathered from 41 volunteer raters. The results of the data were factor analyzed and yielded a two-factor structure including rhythmic support/drive and style/clarity. The 16-item scale accounted for 79.23% of the variance and the alpha reliability was estimated at 0.986. The rhythmic support/drive factor contributed most to discriminating between overall group differences. More specifically, five out of the 16 items contributed most to discriminating between groups.

Keywords
Assessment, facet-factorial, jazz, rating scale, rhythm section

Introduction

The international legitimization of jazz as an artistic music began in the prewar years of the 20th century, notably marked by the overseas performances of traveling American jazz musicians along

Corresponding author:
Brian C. Wesolowski, The University of Georgia, 250 River Road, Athens, GA 30602–7287, USA.
Email: bwes@uga.edu

with the aid of internationally networked media sources such as the radio (Giddins & DeVeaux, 2009; Gioia, 1997; Shipton, 2007; Tirro, 1993). Atkins (2003) writes:

> … practically from its inception, jazz was a harbinger of what we now call 'globalization.' In no one's mind have the music's ties to its country of origin been severed, yet the historical record proves that it has for some time had global relevance. (p. *xv*)

More profound evidence of this globalization lies in the development of an "international jazz culture" that began to emerge in the late 20th century (Atkins, 2003; Baraka, 2009). In particular, Atkins notes, "people around the world have been actively constructing their own systems for performing, understanding, evaluating, and discussing jazz" (2003, p. *xx*). Both Atkins and Shipton (2007) have documented in detail the outpouring of these uniquely developed, localized jazz scenes throughout Europe (i.e., Britain, Germany, France, Sweden, Denmark), South America (i.e., Argentina, Brazil) and Africa (i.e., Zimbabwe, South Africa), as well as in China, Cuba, Canada, and Russia.

Closely linked to the development of this international jazz culture is the global institutionalization of jazz and outgrowth of worldwide jazz education programs (Gonda, 1983). These programs were often pioneered as a result of the "global jazz star network", where international jazz performers developed their careers in New York or other major United States cities, eventually bringing back their experiences and expertise to their native countries (McGee, 2011). Jazz as an international, academic field of study emerged strongly around the turn of the century in both secondary and post-secondary institutions when the International Association for Jazz Education (IAJE) reached its peak of influence (Prouty, 2012). The growth and development of international jazz education programs reiterates the global significance of jazz music, as supported by the sociological notion that education is a reflection of what is valued in a culture (Apple, Ball, & Gandin, 2010; Ballantine & Hammack, 2009; Ballantine & Spade, 2012).

The purpose of this study was to develop a valid and reliable rating scale to assess jazz rhythm sections in the context of jazz big band performance. The research questions that guided this study included:

1. What central factors contribute to the assessment of a jazz rhythm section?
2. What items should be used to describe and assess a jazz rhythm section performance?
3. How should the items be categorized?
4. What differences among jazz rhythm sections exist at three performance achievement levels?
5. What criteria best predict group membership into three performance achievement levels?

Assessing jazz and the rhythm section: Importance and complexities

Higher education institutions seldom offer enough teaching, learning, or performing experiences to adequately prepare pre-service music educators to successfully teach jazz at the secondary school level (Ellis, 1991; Prouty, 2012; Treinen, 2011; West, 2011). According to Coggiola (2004), increased instrumental jazz ensemble experience leads to a greater aesthetic interest in and understanding of complex jazz performances. Evidence also exists that formal musical training and years of experience lead to increased interpretation of pitch, harmony, rhythm, meter, and style (DeCarbo, 1984; Palmer & Krumhansl, 1987a, 1987b, 1990). However, the aural/oral nature of how jazz is taught and learned often escapes undergraduate music education curricula

(Baker, 1989). Aural imitative ability has been identified as a significant predictor of jazz performance (Ciorba, 2009; Greennagel, 1994; Madura, 1996; May, 2003; Watson, 2008). Without this aural development and understanding of its importance in jazz teaching and learning, the rich subtleties of sound, articulation, and stylistic inflection are lost in instruction (Baker, 1989). The need for more specialized and detailed jazz-related instruction in teacher preparation programs is vital to the improvement of jazz pedagogy, particularly when pre-service instrumental music teachers receive the majority of their pedagogical content knowledge from teaching observations, pre-service teaching experience, and course content at the undergraduate level (Haston & Leon-Guerrero, 2008).

Additionally, the unique nature of the rhythm section often inhibits and intimidates music instructors (Dunscomb & Hill, 2002; Lawn, 1995; Thomas, 2008). The rhythm section is considered to be one of the most important sections in the jazz big band (Dunscomb & Hill, 2002). According to Berry (1990), "without a good rhythm section it is nearly impossible to have a good band" (pp. 10–11). The success of a jazz big band relies in part on the performance achievement of its rhythm section (Lawn, 1995). Kuzmich and Bash (1984) indicate that the rhythm section "must be the most well-rehearsed section – the most capable, dependable, and confident section to insure stability for the entire jazz ensemble. Too frequently, however, it is the weakest and most misunderstood section" (p. 155). In a content analysis of high school jazz band festival adjudicator comments, Ellis (1991) found that rhythm sections received a 35% larger proportion of negative comments than other ensemble components. It was concluded that this significantly higher proportion was attributed to five specific difficulties faced by rhythm section players in secondary school ensembles: (a) lack of specified musical notation in the students' parts, (b) lack of specialized instruction and training of the director, (c) difficulty finding qualified private instruction specialists, (d) the self-taught nature of rhythm section students, and (e) that the aural models in popular music (i.e., timbre, technique, and intent) vary greatly from what is found in jazz (Ellis, 2007, pp. 45–46).

Goins (2003) indicates three distinctive characteristics of student learning within the rhythm section: (a) rhythm sections are generally more accustomed to using their ears from the beginning of instruction, (b) their capacity to memorize their music is greater because of the aural-based nature of their musical development, and (c) they are the only members of the ensemble playing their particular instrument, causing them to depend on each other as a cohesive unit (p. 32). Each member of the rhythm section works independently, yet must maintain coherency within the framework of the unit; therefore, the members of the rhythm section are individually and collectively responsible for responding to and accompanying the ensemble (Coker, 1978; Lawn, 1995). In particular, the rhythm section is at the core of the ensemble's vitality and drive. According to Berg, Fischer, Hamilton, and Houghton (2006), "It is the intensity of the rhythm section that 'infects' the remainder of the ensemble, helping them to play with the style and intent of the music" (p. 3).

According to Thomas (2008), "While it is important to focus on the role of each instrument, it is equally important to focus on the role of the rhythm section as a whole" (p. 47). Relative to performance assessment, the accomplishment of individuals, to a large extent, determines the overall success of a group (Jaques, 2000). However, almost all assessment in music education is authentic, where the success of the group is focused on more frequently than the success of the individual (Colwell, 2002, p. 1129; Johnson & Johnson, 2004, p. 86). More importantly, assessing the quality of a group's work is a summative action (Johnson & Johnson, 2004). According to McDonald (1992), a key strategy for planning authentic, summative assessments includes "planning backwards". Specifically, in a musical setting, the instructor would construct an instructional sequence and formative assessments after the consideration of what "real-life" objectives are to be met.

According to Harlen (2012), procedures for ensuring dependable summative assessments "will benefit the formative use, the teacher's understanding of the learning goals, the types of learning tasks that are needed and the nature of progression in achieving them" (p. 100). There are, however, three basic complexities with music performance assessment: (a) the nature of music is subjective due to the encouragement of expressiveness and divergence of response (Radocy, 1986; Wesolowski, 2012), (b) lack of agreement on assessment criteria complicates the assessment process (Bergee, 2003; Fiske, 1983), and (c) many music assessment tools are subject to poor validity and reliability due to weak or inappropriate methodology (Lehman, 2007). The purpose of this study was to develop a valid and reliable summative rating scale to assess the jazz rhythm section using a facet-factorial approach to scale construction. Facet-factorial rating scales have been developed for clarinet performance (Abeles, 1973), euphonium and tuba performance (Bergee, 1987), jazz guitar improvisation (Horowitz, 1994), snare drum performance (Nichols, 1985), guitar performance (Russell, 2010), string performance (Zdzinski & Barnes, 2002), choral performance (Cooksey, 1977), high school band performance (DCamp, 1980), orchestra performance (Smith & Barnes, 2007), and jazz big band performance (Wesolowski, 2015).

The construction of a valid and reliable rating scale for the assessment of a jazz rhythm section based upon valid and reliable testing: (a) provides an opportunity to identify and organize evaluative criteria specific to jazz rhythm section performance, (b) offers a measurement tool for the improvement of teaching, learning, and program accountability, and (c) develops the groundwork for understanding instructional sequences and developing student learning objectives. The evaluation of jazz rhythm sections in the context of a jazz big band was addressed for the reason that most secondary school educational institutions teach the genre of jazz through the medium of a jazz big band. More specifically, most educational institutions in the United States have at least one large jazz ensemble (Dunscomb & Hill, 2002).

Method

Initial item pool generation and measure development

This research study utilized the facet-factorial approach to scale development (Butt & Fiske, 1968, 1969). Items were gathered from research and literature related to the assessment and teaching of jazz rhythm sections (Berg et al., 2006; Berry, 1990; Coker, 1978, Dunscomb & Hill, 2002; Goins, 2003; Jarvis & Beach, 2002; Kernfeld, 1995; Kuzmich & Bash, 1984; LaPorta, 1965; Lawn, 1995; Miles & Carter, 2008; Sherman, 1976; Wheaton, 1975; Wiskirchen, 1966). The relevant items were extracted and altered into statements apt for assessing jazz rhythm sections.

The researcher and three high-school music educators adept in jazz pedagogy screened the item pool. The three high-school music educators have demonstrated long-standing traditions of success, receiving continuous superior ratings and accolades with their jazz ensembles at district, state, and national levels. Items marked as unclear, redundant, or reflecting visual components of a performance were removed. The remaining items ($N = 37$) were grouped into six a priori categories: (a) rhythmic support, (b) rhythmic drive, (c) clarity, (d) balance, (e) musical style, and (f) individual accountability. The a priori categories were based upon researcher intuition and teaching experience. Four collegiate jazz instructors (applied jazz guitar, applied jazz piano, applied jazz bass, and applied jazz drums) were solicited in order to evaluate the item pool. These instructors are full-time faculty members at internationally recognized jazz institutions and each have over 15 years of collegiate teaching experience as well as international performance experience at the highest professional level. Each instructor was asked to independently (a) label the criteria as having a neutral, positive, or negative connotation (i.e., "performs with a stiff swing feel" indicates a negative

attribute for a jazz performance and "maintains a steady tempo" indicates a positive attribute for a jazz performance), (b) verify the aural nature of the item description (i.e., "performs with good posture" indicates a visual component to a performance and would therefore be expunged), and (c) further screen the item pool for redundancy and clarity. Any items that did not receive 100% panel agreement, were considered redundant, or which were labeled as "neutral" were expunged. The inclusion of positively and negatively nuanced criteria items was significant in order to impede the development of specific response sets and to avoid bias (Spector, 1992).

Each statement included in the final item pool ($N = 29$) was paired with a four-point Likert scale. Response alternatives included "Strongly Agree", "Agree", "Disagree", and "Strongly Disagree". A four-point scale was specifically chosen in order to eliminate a neutral category. The elimination of a neutral category provided a better measure of the intensity of participants' attitudes and opinions (Dumas, 1999; Wright, 1977).

The item pool was piloted using 12 jazz big band performances of varying ability levels. The recordings for the pilot study and the full-scale study were solicited from middle-school, high-school, and college educators as well as from professional musicians' personal archives of performances. The ensembles were recorded in a wide range of performance situations, including district and state evaluations, master classes, jazz festivals, and concerts. The researcher carefully screened all recordings in order to control for recording quality and clarity of individual rhythm section instruments. Big band performances were chosen as the aural subjects for the reason that jazz big bands are the mode in which most educational institutions at the secondary-school level teach jazz-related music (Dunscomb & Hill, 2002). The judging panel ($N = 3$) included full-time jazz studies faculty members who direct a jazz big band as part of their course load. The judging panel represented two midsized universities and one large university. All three raters direct a band consisting of undergraduate and graduate music students. The panel was instructed to listen to each of the 12 the performances as many times as needed and to judge the ensemble using the specified measurement tool. Initial data were analyzed and feedback regarding the measurement tool was solicited from the panel. The item pool was revised and edited based upon panelist suggestions. The raters and recordings utilized in the pilot study were eliminated from the full-scale study.

Raters

Volunteer raters ($N = 41$) were solicited based upon their jazz education experience, jazz performance experience, academic background, and availability. The raters were drawn from a pool of musicians and educators specifically skilled in jazz-related music. The pool included secondary-school teachers ($n = 14$), college professors ($n = 8$), graduate jazz performance majors ($n = 6$), graduate music education majors ($n = 4$), and professional musicians ($n = 9$). Of the secondary-school teachers and professional musicians, all have acquired a master's degree in music. Prior to the evaluation, each rater was provided with an instructional packet via email that outlined the premise of the research and clearly defined the task expectations and dimensions of the measurement tool (Winter, 1993). Included in the instructional packet were two anchor recordings for the raters to listen to (Smith, 2009). The anchor recordings provided a reference point that spanned the range of the provided measurement tool. The first recording served to demonstrate a weak performance that would merit a sum score of the lowest possible rating on the measurement tool. The second recording served to demonstrate a strong performance that would merit a sum score of the highest possible rating on the measurement tool. A total of 30 recordings were used in this study. Each rater was given three anonymous recordings from the item pool. The recordings were distributed randomly using an electronic integer generator. The subject of each aural evaluation was a jazz big band with standardized instrumentation (i.e., saxophone section, trumpet section, trombone

section, and rhythm section) selected from a pool of non-commercially-released middle-school, high-school, college, and professional performances. Specifically, each ensemble's rhythm section included an acoustic piano, guitar, bass, and drums. Each musical selection was performed in a medium-tempo, swing style utilizing functional harmony in order to control for style. In particular, the recordings included the full performance of the musical example and ranged between 2 minutes 26 seconds and 5 minutes 38 seconds in length.

Results

A total of 122 responses were collected. One response set was eliminated due to missing data. In order to check for the appropriateness of running a factor analysis, several assumptions were considered. First, all 29 of the evaluative criteria correlated at 0.30 with at least one other item (Zillmer & Vuz, 1995). Second, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.92, above the recommended value of 0.50 (Zillmer & Vuz, 1995). Third, the Bartlett's test of sphericity was significant ($\chi^2_{(406)} = 5583.94$, $p < 0.001$) . Fourth, the communalities were all above 0.30, further confirming that each item shared an appropriate amount of common variance with other items (Zillmer & Vuz, 1995). Last, the sample size was in excess of 100 and the subject-to-variable ratio was above the minimum of 3:1 (Asmus, 1989; Kerlinger & Pedhazur, 1973). In this case, the subject-to-variable ratio was 4.2:1. Given these indicators, factor analysis was found as a suitable method of analysis and was conducted on the 29 items. A reliability analysis was conducted using Cronbach's alpha. The alpha reliability was estimated at 0.986.

Principal axis factoring (PAF) was used in order to account for the variance in common with the variables. PAF was specifically chosen because of the similarity between items and its previous application to aid in the evaluation of musical performances (Asmus & Radocy, 1992, pp. 161–162). The initial eigenvalues showed that the first factor explained 73.19% of the variance, and the second factor 4.52% of the variance. The remaining items reflected eigenvalues under 1.00. One-, two-, three-, and four-factor solutions were examined using oblimin rotations of the factor-loading matrix. Oblimin oblique rotations were specifically utilized because of the nature of relationships between variables (Asmus & Radocy, 1992). The two-factor solution was preferred due to the leveling off of eigenvalues after two factors on the scree plot and the difficulty of interpreting any subsequent factors beyond two. The two-factor solution revealed some items that cross-loaded on both dimensions and other items that did not load highly in either category. The cross-loading and low-loading items were removed from the analysis and re-run in the same manner. The result was a two-factor simple solution containing 16 items, explaining 79.23% of the variance. The factor matrix for this solution is provided in Table 1.

Internal consistency for the Jazz Rhythm Section Performance Rating Scale (JRSPRS) in addition to each of the two factors was examined using Cronbach's alpha. The alpha reliability for the JRSPRS (16 items) was estimated at 0.97. The alpha reliability for Factor 1 (rhythmic support and drive) was 0.98 and the alpha reliability for Factor 2 (style and clarity) was 0.94. No substantial increases in alpha reliability for any of the factors were achieved by eliminating more items.

Scores for each of the 122 Likert scale responses were summed and classified into three performance qualities based upon z probability: upper group (top 25%), middle group (middle 50%), and low group (bottom 25%). An examination of the grouping variable frequency counts adequately supported the classification percentages. Furthermore, composite factor scores were created. Higher scores indicated a stronger performance of the evaluated rhythm sections and lower scores indicated weaker performances of the evaluated rhythm sections. Descriptive and predictive discriminant function analyses were conducted in order to examine and assess the differences between the rhythm sections categorized into each of the performance achievement levels and to test the

Table 1. Factor loadings and communalities based on a principal axis factoring analysis with oblimin oblique rotations for 16 items from the Jazz Rhythm Section Performance Rating Scale (JRSPRS) (*N* = 122).

JRSPRS items	Rhythmic support/drive	Style/clarity	Communality
Rhythm section plays with good support	1.033		0.849
Rhythm section reinforces rhythmic elements of the piece	1.002		0.886
Rhythm section provides a strong, fundamental pulse	0.968		0.921
Rhythm section plays with rhythmic drive	0.937		0.831
Rhythm section maintains a steady tempo	0.887		0.842
Rhythm section plays with rhythmic excitement	0.811		0.877
Rhythm section demonstrates awareness of ensemble parts	0.755		0.800
Fills do not interrupt the steady flow time	0.721		0.777
Rhythm section is rhythmically unified	0.620		0.691
Rhythm section plays with appropriate rhythmic style		1.000	0.808
Chords are voiced appropriately		0.850	0.423
Comping is distributed appropriately		0.845	0.712
Rhythm section interprets their parts appropriately		0.734	0.779
Comping demonstrates rhythmic clarity		0.717	0.692
Rhythm section shows dynamic sensitivity		0.650	0.730
Comping is done in appropriate registers		0.625	0.659

Note: Factor loadings < 0.3 are suppressed.

Table 2. Descriptive statistics for the two factor score items, two-group Jazz Rhythm Section Performance Rating Scale data.

		Group means (SDs)			Univariate
		High	Middle	Low	*F* (*p*)
Rhythmic support/drive	(Y1)	32.65 (0.41)	23.34 (0.67)	13.61 (0.54)	320.71 (*p* < 0.001)
Style/clarity	(Y2)	24.72 (0.40)	16.46 (0.49)	11.34 (0.33)	2.05 (*p* = 0.14)

Note: Shaded areas indicate construct identifiers.

predictability of the two derived factors (i.e., rhythmic support/drive and style/clarity). The sample was split into two portions. A 65% analysis sample and a 35% holdout sample were applied. The three criterion groups were defined as high (*n* = 30), middle (*n* = 25), and low (*n* = 20). No missing data were found in the 122 x 16 data matrix. Descriptive information for the two factor scores and 16 items is provided in Tables 2 and 3.

According to the univariate tests, Rhythmic support/drive (Y_1) contributed most to the overall group differences. More specifically, "Rhythm section demonstrates awareness of ensemble parts" (X_5), "Rhythm section reinforces rhythmic elements of the piece" (X_6), "Fills do not interrupt the steady flow time" (X_8), "Chords are voiced appropriately" (X_{12}) and "Comping demonstrates rhythmic clarity" (X_{15}) were the variables contributing most to the overall group differences.

Linear discriminant functions (LDFs) were considered to further examine the group differences. One statistically significant discriminant function for the two factor scores and two statistically significant discriminant functions for the 16 items resulted (See Table 4).

Table 3. Descriptive statistics for the 16 items, two-group Jazz Rhythm Section Performance Rating Scale data.

		Group means (SDs)			Univariate
		High	Middle	Low	F *
Rhythmic support/drive	(Y$_1$)				
Rhythm section plays with good support	(X1)	3.70 (0.47)	2.68 (0.63)	1.50 (0.51)	101.38
Rhythm section plays with rhythmic excitement	(X2)	3.70 (0.47)	2.56 (0.65)	1.70 (0.57)	79.32
Rhythm section provides a strong, fundamental pulse	(X3)	3.73 (0.45)	2.52 (0.65)	1.60 (0.50)	97.76
Rhythm section plays with rhythmic drive	(X4)	3.33 (0.48)	2.68 (0.56)	1.60 (0.50)	68.70
Rhythm section demonstrates awareness of ensemble parts	(X5)	3.87 (0.35)	2.60 (0.71)	1.50 (0.61)	110.44
Rhythm section reinforces rhythmic elements of the piece	(X6)	3.50 (0.51)	2.60 (0.50)	1.60 (0.50)	85.86
Rhythm section maintains a steady tempo	(X7)	3.80 (0.41)	2.68 (0.75)	1.65 (0.49)	89.40
Fills do not interrupt the steady flow time	(X8)	3.43 (0.50)	2.80 (0.65)	1.30 (0.57)	84.66
Rhythm section is rhythmically unified	(X9)	3.43 (0.63)	2.28 (0.68)	1.65 (0.59)	51.46
Style/clarity	(Y$_2$)				
Rhythm section plays with appropriate rhythmic style	(X10)	3.60 (0.68)	2.56 (0.65)	1.75 (0.55)	52.65
Comping is distributed appropriately	(X11)	3.20 (1.03)	1.64 (0.70)	1.65 (0.49)	33.29
Chords are voiced appropriately	(X12)	3.57 (0.68)	2.24 (0.52)	1.55 (0.51)	77.15
Rhythm section interprets their parts appropriately	(X13)	3.63 (0.56)	2.72 (0.74)	1.60 (0.60)	62.21
Rhythm section shows dynamic sensitivity	(X14)	3.57 (0.63)	2.36 (0.76)	1.80 (0.41)	52.89
Comping demonstrates rhythmic clarity	(X15)	3.57 (0.50)	2.64 (0.76)	1.40 (0.50)	78.30
Comping is done in appropriate registers	(X16)	3.43 (0.50)	2.52 (0.71)	1.45 (0.61)	64.31

Note: Shaded areas indicate construct identifiers.
*df_1 = 2, df_2 = 72, p < 0.001.

Table 4. Test of dimensionality for the three-group Likert scale response data.

Number of dimensions	Λ	F	$Df_{b/w}$, $Df_{w/i}$	p
1	0.04	24.37	32, 208	< 0.001
2	0.58	5.08	15, 105	< 0.001

Table 5. Variable ordering for the three-group Likert scale response data.

Variable	*F* to remove
Rhythm section demonstrates awareness of ensemble parts (*X5*)	10.20
Comping demonstrates rhythmic clarity (*X15*)	6.78
Rhythm section reinforces rhythmic elements of the piece (*X6*)	5.16
Chords are voiced appropriately (*X12*)	15.86
Fills do not interrupt the steady flow of time (*X8*)	6.21

Examination of the group means indicated that the high achieving rhythm sections were separated out from the middle and low achieving groups using the two factor scores. With respect to the 16 items, it appeared that the high achieving group was separated from the middle and low achieving groups on function one and the middle achieving group was separated from the high and low achieving groups on function two. Using a stepwise method for variable selection, the model indicated that when considering the two factor scores, Factor 1 (rhythmic support/drive) was the most important factor in discriminating between the high achieving group from the other two groups. When considering the 16 individual items, the model indicated that "Rhythm section demonstrates awareness of ensemble parts" (X_5), "Rhythm section reinforces rhythmic elements of the piece" (X_6), and "Chords are voiced appropriately" (X_{12}) maximally differentiated the high and middle achieving groups. "Fills do not interrupt the steady flow time" (X_8) and "Comping demonstrates rhythmic clarity" (X_{15}) maximally differentiated the middle and low achieving groups. The variable ordering is listed in Table 5.

The cross-validation (leave-one-out) rule was reported because an external classification rule is recommended to estimate group hit rates (Huberty & Olejnik, 2006, p. 325). The predictive accuracy for the cross-validation sample was 94.7% with consideration towards the two factor scores and 85.3% with consideration towards the 16 items. The predictive accuracy for the holdout sample was 91.5% for the two factor scores and 89.4% for the 16 items. The proportional chance criterion for assessing model fit was 0.69. Based upon the requirement that model accuracy be 25% better than chance criteria, the model accuracy of the holdout sample of the 16-item examination exceeded the standard by 3.3% (Hair, Anderson, Tatham, & Black, 1998). The 91.5% and 89.4% classification accuracies in the holdout sample provide additional evidence of the predictive validity of the model (Hair et al., 1998). Tables containing predictive discriminant function analysis data are available upon request.

Provided the factor score for rhythmic support/drive or the scores of the predictor variables X_5, X_6, X_8, X_{12}, and X_{15} are available, a composite score can be obtained by (a) multiplying each predictor score by the respective weight, (b) summing the remaining products, and (c) adding the constant (Huberty & Olejnik, 2006, p. 325). Once a composite score is calculated, the newly assessed rhythm section can be assigned to the achievement level for which the largest composite score is obtained.

Conclusion and discussion

The purpose of this study was to develop a valid and reliable rating scale to assess jazz rhythm sections in the context of jazz big band performance. The factor analysis revealed that there were two distinct factors that contribute to the assessment of a jazz rhythm section: rhythmic support/drive and style/clarity. Factor 1 (i.e., rhythmic support/drive) consisted of nine distinct items and Factor 2 (i.e., style/clarity) consisted of seven distinct items (see Table 1). Upon splitting the assessed jazz rhythm section performances into three achievement levels (i.e., upper, top 25%;

middle, middle 50%; and low, bottom 25%), it was found that the rhythmic support/drive factor contributed the most to the overall group differences. However, the specific items that contributed most to the group differences included "Rhythm section demonstrates awareness of ensemble parts" (X_5), "Rhythm section reinforces rhythmic elements of the piece" (X_6), "Fills do not interrupt the steady flow time" (X_8), "Chords are voiced appropriately" (X_{12}) and "Comping demonstrates rhythmic clarity" (X_{15}). The reliability of the JRSPRS was estimated at 0.97.

The development of the JRRPRS's classification rule allows for newly assessed rhythm sections to be assigned to a high, middle or low achievement level with an 85.3% predictability accuracy when utilizing the 16 individual items. The benefit of this rule is twofold. First, its implementation as a formal performance assessment may enlighten students and directors as to their achievement level on the normal curve as well as the specific construct identifiers to be discussed for future improvement. Second, its application as an initial diagnostic tool may provide a platform for the director and students as to their current performance achievement level. In turn, this offers concrete and tangible student learning objectives that can guide the rehearsal and teaching processes. According to Maki (2010), planning the assessment process backwards (i.e., planning formative assessments and instructional strategies based upon a pre-existing summative assessment tool):

> … slows us [instructors] down to think about not only what we want to assess, but also whom we want to assess, how we want to assess, and when we want to derive evidence of students' enduring learning along their educational journey. Without raising and answering questions about our students and what and when we want to learn about their achievements – before launching into the assessment process – our results may not be all that useful or may lend themselves to limited generalizations … (p. 7).

The development of this assessment tool also sheds light on learning sequences within the jazz rhythm section. Considering solely the factor scores, it was found that the high achieving groups were separated from, collectively, the middle and low achieving groups on the rhythmic support/drive factor. In considering the 16 items of the scale independently, it was found that that "Rhythm section demonstrates awareness of ensemble parts" (X_5), "Rhythm section reinforces rhythmic elements of the piece" (X_6), and "Chords are voiced appropriately" (X_{12}) maximally differentiated the high and middle achieving groups. "Fills do not interrupt the steady flow time" (X_8) and "Comping demonstrates rhythmic clarity" (X_{15}) maximally differentiated the middle and low achieving groups. An understanding of which items differentiated the high, middle, and low achieving groups demonstrates specific, measurable areas of student growth and development at multiple achievement levels. This information provides reliable data to use as a foundation for constructing learning sequences in jazz performance.

In order to develop a more comprehensive assessment program for secondary-school jazz rhythm section performance, it is suggested that a formative assessment tool based upon the JRSPRS be developed and validated. A formative assessment tool, such as a criteria-specific analytic rubric, may provide a more comprehensive method for the facilitation of learning and teaching as well as offer a mode for directly engaging students in the learning process. The combination of a formative assessment tool with the rating scale developed through this research may aid in the development of a framework for assessing members of a jazz rhythm section by providing a comprehensive and accurate account of student learning in the classroom. Additionally, the implementation of concrete measurement tools in the instructional process provides a clearer connection between teacher expectation and student learning by providing a detailed description of instructional expectations, therefore improving the instructional process.

The process of rater-mediated scale construction is always bound with inaccuracy as a result of human conditions. The challenge of measuring complex behaviors, comprehensively defining

measurement constructs, and overcoming the nuance of scoring makes performance assessment a complex task and provides grounds for plausible validity arguments. Two particular areas of further research can benefit music performance assessment. First, little work has been done in music education with the application of modern test theory (i.e., item response theory models) to music performance assessment in a rater capacity. Second, research on rater accuracy may prove to be fruitful in improving validity and reliability in scale construction. Both areas can be investigated through the application of a many-facet Rasch model (i.e., invariant measurement) to the raw data gleaned from this study. The scale development process could be enlightened by standardizing the raw data on a logit scale, identifying underlying latent variable constructs, identifying leniency and severity of raters (i.e., rater accuracy), inferring true scoring, and indicating item difficulty. The application of such would better illustrate the relationship between judgmental processes and model-based approaches to measurement.

Funding

References

Abeles, H. A. (1973). Development and validation of a clarinet performance adjudication scale. *Journal of Research in Music Education*, *21*(3), 246–255.

Apple, M. W., Ball, S. J., & Gandin, L. A. (2010). *The Routledge international handbook of the sociology of education*. London, UK: Routledge.

Asmus, E. P. (1989). Factor analysis: A look at the technique through the data of Rainbow. *Bulletin of the Council for Research in Music Education*, *85*, 1–13.

Asmus, E. P., & Radocy, R. E. (1992). Quantitative analysis. In R. Colwell (Ed.), *Handbook of research of music teaching and learning* (pp. 141–195). New York, NY: Schirmer Books.

Atkins, E. T. (Ed.). (2003). *Jazz planet*. Jackson: University Press of Mississippi.

Baker, D. (1989). *Jazz pedagogy: A comprehensive method of jazz education for teacher and student*. Van Nuys, CA: Alfred Publishing.

Ballantine, J. H., & Hammack, F. M. (2009). *The sociology of education: A systematic analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.

Ballantine, J. H., & Spade, J. Z. (2012). *Schools and society: A sociological approach to education*. Los Angeles, CA: Sage/Pine Forge Press.

Baraka, A. (2009). *Digging: The Afro-American soul of American classical music*. Berkeley: University of California Press.

Berg, S., Fischer, L., Hamilton, F., & Houghton, S. (2006). *Rhythm section workshop for jazz directors*. Van Nuys, CA: Alfred Publishing.

Bergee, M. J. (1987). *An application of the facet-factorial approach to scale construction in the development of a rating scale for euphonium and tuba music performance* (Unpublished doctoral dissertation). University of Kansas, USA.

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

Berry, J. (1990). *The jazz ensemble director's handbook: Understanding the A-to-Zs of each section, including the rhythm section*. Milwaukee, WI: Jenson Publications.

Butt, D. C., & Fiske, D. W. (1968). A comparison of strategies in developing scales for dominance. *Psychological Bulletin*, *70*, 505–519.

Butt, D. C., & Fiske, D. W. (1969). Differential correlates of dominance scales. *Journal of Personality*, *37*(3), 415–428.

Ciorba, C. R. (2009). Predicting jazz improvisation achievement through the creation of a path analytical model. *Bulletin of the Council of Research in Music Education*, *180*, 43–57.

Coggiola, J. C. (2004). The effect of conceptual advancement in jazz music selections and jazz experience on musicians' aesthetic response. *Journal of Research in Music Education*, *52*(1), 29–42.

Coker, J. (1978). *Listening to jazz*. Englewood Cliffs, NJ: Prentice Hall.

Colwell, R. (2002). Assessment's potential in music education. In R. Colwell & C. Richardson (Eds.), *The new handbook of research on music teaching and learning* (pp. 1128–1156). Oxford, UK: Oxford University Press.

Cooksey, J. M. (1977). A facet-factorial approach to rating high school choral music performance. *Journal of Research in Music Education*, *25*, 100–114.

DCamp, C. B. (1980). *An application of the facet-factorial approach to scale construction in the development of a rating scale for high school band performance* (Doctoral dissertation). University of Iowa, USA. (*Dissertation Abstracts International, 41*, 1462A.)

DeCarbo, N. (1984). The effects of years of teaching experience and major performance instrument on error detection scores of instrumental music teachers. *Contributions to Music Education*, *46*(2), 182–192.

Dumas, J. (1999). *Usability testing methods: Subjective measures, part II – Measuring attitudes and opinions*. Washington, DC: American Institutes for Research.

Dunscomb, J. R., & Hill, W. (2002). *Jazz pedagogy: The jazz educator's handbook and resource guide*. Miami, FL: Warner Bros. Publications.

Ellis, M. C. (1991). An analysis of taped comments from a high school jazz band festival. *Contributions to Music Education*, *34*, 35–49.

Fiske, H. E. (1983). Judging music performance: Method or madness? *Update: The Applications of Research in Music Education*, *1*(3), 7–10.

Giddins, G., & DeVeaux, S. K. (2009). *Jazz*. New York, NY: W. W. Norton.

Gioia, T. (1997). *The history of jazz*. New York, NY: Oxford University Press.

Goins, W. E. (2003). *The jazz band director's handbook: A guide to success*. Lewiston, NY: Edwin Mellen Press.

Gonda, J. (1983). Jazz education: Improvisation and creativity. *International Journal of Music Education*, *2*, 19–22.

Greennagel, D. J. (1994). *A study of selected predictors of jazz vocal improvisation skills* (Doctoral dissertation). Retrieved from http://search.proquest.com/pqdt/index.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 87–102). Thousand Oaks, CA: Sage.

Haston, W., & Leon-Guerrero, A. (2008). Sources of pedagogical content knowledge: Reports by preservice instrumental music teachers. *Journal of Music Teacher Education*, *17*(2), 48–59.

Horowitz, R. A. (1994). *The development of a rating scale for jazz guitar improvisation performance* (Unpublished doctoral dissertation). Columbia University Teachers College, USA. (*Dissertation Abstracts International, 55*(11A), 3443.)

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). Hoboken, NJ: Wiley InterScience.

Jaques, D. (2000). *Learning in groups: A handbook for improving group work*. Sterling, VA: Stylus Publishing.

Jarvis, J., & Beach, D. (2002). *The jazz educator's handbook*. Delevan, NY: Kendor Music.

Johnson, D. W., & Johnson, R. T. (2004). *Assessing students in groups: Promoting group responsibility and individual accountability*. Thousand Oaks, CA: Corwin Press.

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York, NY: Holt, Rinehart, and Winston.

Kernfeld, B. (1995). *What to listen for in jazz*. New Haven, CT: Yale University Press.

Kuzmich, J., & Bash, L. (1984). *Complete guide to instrumental jazz instruction: Techniques for developing a successful jazz program*. West Nyack, NY: Parker Publishing.

LaPorta, J. (1965). *Developing the school jazz ensemble*. Boston, MA: Berklee Press.

Lawn, R. (1995). *The jazz ensemble director's manual: A handbook of practical methods and materials for the educator*. Oskaloosa, IA: C. L. Barnhouse.

Lehman, P. R. (2007). Getting down to basics. In T. S. Brophy (Ed.), *Assessment in music education: Integrating curriculum, theory, and practice* (pp. 17–27). Chicago, IL: GIA Publications.

Madura, P. D. (1996). Relationships among vocal jazz improvisation achievement, jazz theory knowledge, imitative ability, musical experience, creativity, and gender. *Journal of Research in Music Education*, *44*, 252–267.

Maki, P. L. (2010). *Assessing for learning*. Sterling, VA: Stylus Publishing.

May, L. F. (2003). Factors and abilities influencing achievement in instrumental jazz improvisation. *Journal of Research in Music Education*, *51*(3), 245–258.

McDonald, J. P. (1992). Dilemmas of planning backwards: Rescuing a good idea. *Teachers College Record*, *94*(1), 152–169.

McGee, K. (2011). New York comes to Groningen: Jazz star circuits in the Netherlands. In J. Toynbee & B. Dueck (Eds.), *Migrating music* (pp. 202–217). London, UK: Routledge.

Miles, R., & Carter, R. (Eds.). (2008). *Teaching music through performance in jazz*. Chicago, IL: GIA Publications.

Nichols, J. P. (1985). *A factor analysis approach to the development of a rating scale for snare drum performance* (Doctoral dissertation). University of Iowa, USA. (*Dissertation Abstracts International, 46*, 3282A.)

Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in perception of musical phrases. *Journal of Experimental Psychology*, *15*, 331–346.

Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, *41*, 505–518.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology*, *16*, 728–741.

Prouty, K. (2012). *Knowing jazz: Community, pedagogy, and canon in the information age*. Jackson: University Press of Mississippi.

Radocy, R. E. (1986). On quantifying the uncountable in musical behavior. *Bulletin of the Council for Research in Music Education*, *88*, 22–31.

Russell, B. E. (2010). The development of a guitar performance rating scale using a facetfactorial approach. *Bulletin of the Council of Research in Music Education*, *184*, 21–34.

Sherman, H. (1976). *Techniques and materials for stage band*. Los Angeles, CA: Creative World Music Publications.

Shipton, A. (2007). *A new history of jazz*. New York, NY: Continuum.

Smith, B. P., & Barnes, G.V. (2007). Development and validation of an orchestra performance rating scale. *Journal of Research in Music Education*, *55*, 268–280.

Smith, D. T. (2009). Development and validation of a rating scale for wind jazz improvisation performance. *Journal of Research in Music Education*, *57*(3), 217–235.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.

Thomas, R. (2008). The rhythm section: The band within the band. In R. Miles & R. Carter (Eds.), *Teaching music through performance in jazz* (pp. 47–61). Chicago, IL: GIA Publications.

Tirro, F. (1993). *Jazz: A history*. New York, NY: Norton.

Treinen, C. M. (2011). *Kansas high school band directors' and college faculties' attitudes towards teacher preparation in jazz education* (Unpublished doctoral dissertation). Kansas State University, USA.

Watson, K. E. (2008). *The effect of aural versus notated instructional materials on achievement and self-efficacy in jazz improvisation* (Doctoral dissertation). Retrieved from http://search.proquest.com/pqdt/index.

Wesolowski, B. C. (2012). Understanding and creating rubrics for the assessment of music performance. *Music Educators Journal*, *98*(36), 36–42.

Wesolowski, B. C. (2015). Assessing jazz big band performance: The development, validation, and application of a facet factorial rating scale. *Psychology of Music*. Advance online publication. doi:10.1177/0305735614567700

West, C. L. (2011). *Teaching middle school jazz: An exploratory sequential mixed methods study*. (Unpublished doctoral dissertation). Kansas State University, USA.

Wheaton, J. (1975). *How to organize and develop the stage band: Director's manual*. North Hollywood, CA: Maggio Music Press.

Winter, N. (1993). Music performance assessment: A study of the effects of training and experience on the criteria used by music examiners. *International Journal of Music Education*, *22*, 34–39.

Wiskirchen, G. (1966). *Developmental techniques for the jazz ensemble musician*. Boston, MA: Berklee Press.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97–116.

Zdzinski, S. F., & Barnes, G. V. (2002). Development and validation of a string performance rating scale. *Journal of Research in Music Education*, *50*, 245–255.

Zillmer, E. A., & Vuz, J. (1995). Factor analysis with Rorschach data. In J. E. Exner Jr. (Ed.), *Methods and issues in Rorschach research* (pp. 251–306). Hillsdale, NJ: Lawrence Erlbaum Associates.