



## MENC: The National Association for Music Education

---

Development and Validation of a Rating Scale for Wind Jazz Improvisation Performance

Author(s): Derek T. Smith

Source: *Journal of Research in Music Education*, Vol. 57, No. 3 (Oct., 2009), pp. 217-235

Published by: [Sage Publications, Inc.](#) on behalf of [MENC: The National Association for Music Education](#)

Stable URL: <http://www.jstor.org/stable/40343728>

Accessed: 19/11/2013 22:48

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Sage Publications, Inc.* and *MENC: The National Association for Music Education* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Research in Music Education*.

<http://www.jstor.org>

# Development and Validation of a Rating Scale for Wind Jazz Improvisation Performance

Journal of Research in

Music Education

57(3) 217–235

© 2009 MENC: The National

Association for Music Education

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/0022429409343549

<http://jrme.sagepub.com>



Derek T. Smith<sup>1</sup>

## Abstract

The purpose of this study was to construct and validate a rating scale for collegiate wind jazz improvisation performance. The 14-item Wind Jazz Improvisation Evaluation Scale (WJIES) was constructed and refined through a facet-rational approach to scale development. Five wind jazz students and one professional jazz educator were asked to record two improvisations accompanied by an Aebersold play-along compact disc. Sixty-three adjudicators evaluated the 12 improvisations using the WJIES and the Instrumental Jazz Improvisation Evaluation Measure. Reliability was good, with alpha values ranging from .87 to .95. Construct validity for the WJIES was confirmed through the analysis of a multitrait-multimethod matrix. The results of this study indicate that the facet-rational approach is an effective method of developing a rating scale for collegiate wind jazz improvisation performance.

## Keywords

improvisation, jazz pedagogy, jazz performance, music rating scale

With its origins dating back to the turn of the 20th century, jazz has progressed from an American music curiosity to an international phenomenon. This popularity undoubtedly has contributed to the increase in jazz instruction within schools and colleges. Over recent years, jazz as an academic discipline has grown in volume and stature. Indeed, jazz studies now play a significant role in a number of music programs within the higher education sector (Whyton, 2006). Improvisation can be

---

<sup>1</sup>Lincoln University of Missouri, Jefferson City, MO, USA

## Corresponding Author:

Derek T. Smith, 5009 Stone Mountain Parkway, Columbia, MO 65201 USA

E-mail: [smithd1@lincolnu.edu](mailto:smithd1@lincolnu.edu)

found in every music style and culture, but in jazz, it is the predominant and driving force (Kynaston & Ricci, 1978). Jazz has brought about a renaissance in improvisation, providing a style conducive to spontaneous creation (Coker, 1964). In January 1994, the National Committee for Standards in the Arts announced America's first national voluntary standards for K–12 education in the arts (MENC, 2001). Standard Three, *Improvising melodies, variations, and accompaniments*, specifically advocates instruction in improvisation as part of a holistic music curriculum.

However, the aggregate of musicians, educators, and researchers has yet to agree upon what elements constitute the core of jazz improvisation. Many believe that spontaneous creation lies at the heart of the improvisational process (Coker, Casale, Campbell, & Greene, 1970; Gridley, 1991). Others, such as Berliner (1994), dispute the term *spontaneous*. He stated, "Performance without previous preparation is fundamentally misleading. There is, in fact, a lifetime of preparation and knowledge behind every idea that an improviser performs" (p. 17). Coker (1964) identified five factors that chiefly determine the outcome of a jazz player's improvisation: intuition, intellect, emotion, sense of pitch, and habit. Kenny and Gellrich (2002) suggested that "two key constraints of improvisation—knowledge bases and referents—work together to generate new music structures" (p. 119). Although music researchers have attempted earnestly to unravel the process of jazz improvisation, research on the subject continues to be sparse in comparison to other areas of music education.

The current literature related to jazz improvisation, although limited, is diverse. In order to better understand the theoretical constructs that generate improvisation, an attempt has been made to model its salient features (Kenny & Gellrich, 2002). Pressing (1987) presented a highly developed cognitive model in which improvisation was construed as an ordered sequence of *event clusters*. Through his computer-based model, Johnson-Laird (1991) suggested that improvisational creativity should be computable and that only three sorts of algorithms can be creative. The essence of this theoretical model is the internalization (long-term memory) and automation (readily accessible through practice and performance) of the knowledge base (previously learned material) (Kenny & Gellrich, 2002).

In addition to defining the theoretical process of improvisation, researchers have sought to understand how certain factors and instructional methodologies influence improvisation achievement. Several conclusions can be drawn from the collective findings. The results of these studies indicated that it is possible to create an objective instrumental measure of improvisation (Burnsed, 1978; May, 2003). Various instructional sequences have been shown to be effective methods of teaching improvisation, including nontechnical, aural, and visual approaches (Aitken, 1975; Bash, 1983; Burnsed, 1978; Damron, 1973; Hores, 1977). Variables such as self-evaluation of improvisation skill and aural imitation were found to be good predictors of jazz improvisation achievement (Greennagel, 1994; May, 2003). In addition to effectively improving improvisation skill, instructional sequences related to improvisation may also result in improved attitudes toward improvisation and the particular instruction method utilized (Berard, 1998).

Instruction in any music performance medium invariably calls for a valid and reliable method of evaluation (Bergee, 2003). Jazz improvisation can be manifested in a virtually infinite number of acceptable musical outcomes. As a result, jazz improvisation evaluation presents additional reliability challenges. However, despite the potential for added subjectivity, several researchers have succeeded in developing reliable and valid improvisation measurements (Horowitz, 1994; May, 2003; McPherson, 1995; Pfenninger, 1990). If reliable and valid instrumental jazz improvisation measures presently exist, is it truly necessary to construct additional rating scales?

Although the existing measures have proved to be reliable and valid, there remains an opportunity for improvement, thus potentially producing an evaluation that more closely reflects the construct of wind jazz improvisation. The measures developed by May (2003) and McPherson (1995) both contain a total of seven items. These small item pools may only partly describe jazz improvisation performance, especially when compared to the 40-item scale created by Pfenninger (1990) and the 30-item scale produced by Horowitz (1994). Pfenninger (1990) divided the improvisation construct into three distinct dimensions: (a) tonal, (b) rhythm, and (c) expression. A separate scale was developed to evaluate each of the three dimensions. Although both the tonal and rhythm scales proved to be reliable measures, the expression scale was shown to be unreliable. Because Horowitz's Jazz Guitar Improvisation Rating Scale (JGIRS; Horowitz, 1994) includes a lack of inherently "guitaristic" items, he suggested further research to determine the suitability of the JGIRS for use with certain instruments. He raised several questions related to the scale's general usability. Could one scale be used for all instruments, or should different scales be developed using some, but not all, common items (Horowitz, 1994)?

Numerous music performance measures have been developed using the facet-factorial approach to scale construction outlined by Butt and Fiske (1968). This method has proved to be an effective means of creating reliable and valid performance measures (Abeles, 1973; Bergee, 1987; Dcamp, 1980; Jones, 1986; Pazitka-Munroe, 2002; Zdzinski, 2002). A few reliable jazz improvisation scales also have been constructed using this technique (Horowitz, 1994; May, 2003). However, research exploring the viability of using the facet-rational approach to scale development, discussed by Butt and Fiske (1968), remains limited. In addition, the use of a multitrait-multimethod (MTMM) matrix outlined by Campbell and Fiske (1959) to assess the construct validity of various music performance measures has been sparse. Despite limited use in the context of music performance evaluation, the MTMM matrix constitutes one of the most advanced means for assessing construct validity (Campbell & Fiske, 1959).

Historically, learning to improvise in the genre of jazz encompassed an individual exploratory process of listening and imitation. Evaluation of improvisation was limited essentially to either praise or admonition by an audience or by one's peers. This methodology invariably raises issues of reliability and validity. Consequentially, the inclusion of jazz studies within the structured environment of the school

curriculum demands that the methods utilized to evaluate jazz improvisation be both valid and reliable. Students must be provided with appropriate feedback in order to improve performance. Likewise, jazz educators must employ measurement tools that adequately measure the intended phenomenon in order to provide appropriate instruction. If music curricula are to offer instruction in jazz, methods of evaluating improvisation must be developed that more accurately account for all the factors that define improvisation. Accordingly, the purpose of this study was to develop a valid and reliable rating scale designed to evaluate the improvisation ability of collegiate wind instrumentalists. The study also was intended to examine the viability of the facet-rational approach to scale development, as well as the feasibility of utilizing a MTMM matrix to assess validity.

## Method

### *Measure Development*

E-mail solicitations were sent out to a variety of jazz musicians, educators, and advanced jazz students. Each was asked to write down descriptors for either a good or poor wind instrumental improvisation. By including a variety of jazz musicians at varying stages of development in the process, a diverse, comprehensive item pool was generated. Additional descriptors were produced by the examination and analysis of interviews of jazz wind instrumentalists published in *Downbeat*, *Jazzed*, the *Jazz Educator's Journal*, and other jazz-related publications including *Saxophone Journal* and *Jazziz*. Pedagogical materials related to the teaching of jazz improvisation also were reviewed and analyzed for additional improvisation descriptors. This material existed in the form of books, guides, or audiovisual instructional aids. The final source for improvisation descriptors came from the research literature. A few improvisation scales have been developed previously that exhibited acceptable levels of both reliability and validity (Horowitz, 1994; May, 2003; McPherson, 1995; Pfenninger, 1990). Relevant items were borrowed from these measures and added to the initial item pool.

Once all of the items were collected, the pool was examined for clarity and redundancy. Redundant items were combined or eliminated. Some items were reworded in order to enhance clarity, while others were discarded altogether because of ambiguity. Each of the 85 items was placed into one of 14 item groupings by the researcher (for additional details, see D. Smith, 2007). Initially, group placement of each item was based on three main considerations: a review of jazz related literature, performance measurement research, and the intuition of the researcher. These groupings represented 14 distinct evaluation areas. The items within each group were combined and/or modified in order to produce one concise descriptor for each of the 14 groupings. The resulting item pool was reviewed by a panel of jazz educators who provided feedback related to the clearness, appropriateness, and conciseness of the items. At the conclusion of the refinement process, remaining items randomly were phrased

positively and negatively to avoid acquiescence bias (Spector, 1992). Each item was then paired with a 7-point Likert-type scale. Responses included *disagree strongly*, *disagree moderately*, *disagree slightly*, *neither agree nor disagree*, *agree slightly*, *agree moderately*, and *agree strongly*. Although 5-point scales have been shown to provide an adequate number of response categories (Abeles, 1973; Bergee, 1987; Horowitz, 1994; Zdzinski, 2002), I chose to use a 7-point scale after a review of the Instrumental Jazz Improvisation Evaluation Measure created by May (2003). I hoped that the addition of more precise degrees of agreement and/or disagreement would produce a more accurate measurement scale capable of successfully discriminating between performances of similar quality.

Prior to finalizing the Wind Jazz Improvisation Evaluation Scale (WJIES), the measure was piloted by two faculty members at a large Midwestern university. Both adjudicators possessed extensive jazz experience, one as a distinguished performer and the other as a performer/educator. The judges were given a packet containing two copies of the WJIES and a compact disc recording of two improvised jazz solos. The adjudicators were asked to evaluate each solo using the WJIES and to provide feedback related to its usability. Feedback was positive. Items were seen to be concise, easy to understand, and relevant. One pilot adjudicator also commented on the need to read each item carefully in order to ascertain whether or not the item was worded negatively or positively. Based on the resulting feedback provided by both evaluators, the wording of one of the items was modified slightly and the 14-item WJIES was then finalized.

This method of scale development is consistent with procedures first discussed by Butt and Fiske (1968). They referred to this methodology as the facet-rational approach to scale development. A facet-rational approach involves the test constructor developing and/or choosing items based on personal rationale or preconceived ideas related to a particular subject. This method also assumes that the construct in question is composed of various facets. A facet-rational approach differs from a facet-factorial approach in that factor analysis is not used initially to determine the item pool. Factor analysis may then be employed as a method of content validation.

## Participants

**Instrumentalists.** Five wind jazz students and one professional jazz educator were asked to record two improvisations. Five of the six performers self-described their improvisation ability as either moderate or advanced. Four of the six performers reported from 1 to 5 years of piano experience. In addition, four of the six participants also reported 1 to 3 hours of jazz listening per week, with the remaining two reporting listening for 4 or more hours. Collegiate improvisers consisted of three juniors, one senior, and one graduate student. The professional musician (trombonist) provided improvisations at a high level of sophistication. The instrumentalists were chosen purposively based on their improvisation ability in an attempt to provide the judges with a varied set of stimuli (two saxophonists, two trumpeters, and

two trombonists). The improvisers were sorted based on improvisation ability into one of three descriptive categories: (a) intermediate, (b) advanced, or (c) sophisticated. I attempted to facilitate the creation of two stimuli from each category. The placement and determination of student improvisation ability was decided by the collaborative observations of the researcher and the jazz studies director at the university. Decisions were based on past improvisation performances and/or successful completion of jazz improvisation course work. Delineation was done in such a way as to create three groups of two. The intermediate category included a trombonist and a trumpeter. The advanced category included a trumpeter and an alto saxophonist. The sophisticated category included a tenor saxophonist and the professional trombonist.

**Adjudication.** According to procedures outlined by Gorsuch (1983), the ideal ratio of adjudicator-to-scale items should be about 5 to 1 when utilizing factor analysis as part of the content validation process. Therefore, the 14 items on the WJIES called for an adjudicator panel of about 70 members. Seventy-one adjudication packets were compiled and presented to potential evaluators. Sixty-three of the adjudicator packets were returned, resulting in an adjudicator-to-scale ratio of 4.5 to 1. Adjudicators consisted of university jazz students enrolled at a large Midwestern university, college jazz students enrolled at a small rural university, and, various jazz educators and experienced jazz performers. A smaller adjudicator panel consisted of the 4 most highly experienced jazz musicians among the original 63. Three of these judges were jazz educators with 10 to 20 years of experience teaching and evaluating jazz improvisation. At least 2 of the evaluators had worked as professional jazz musicians for at least 3 years prior to returning to college. Overall, the evaluators were selected to represent a variety of age groups, improvisational ability, and jazz instruction experience. The intent was that the WJIES could be utilized reliably by a wide spectrum of jazz educators, performers, and students.

In order to generate global evaluation scores, an advanced jazz musician was asked to listen to the 12 recorded improvised solos and rate them based on two traits, performance skills and creative development. This individual possessed 10 years of experience performing instrumental wind jazz in both North American and South American venues. He was excluded purposely from the original adjudicator pool in order to be available to provide the global evaluations. The traits emerged as a result of factor analyzing the 12 performances as measured by the WJIES. A 5-point scale with 5 being the high score was employed. Consistent with the traditional rating schemes for adjudicated festivals, a 5-point scale was deemed appropriate based on the nature and purpose of the global evaluations (B. Smith & Barnes, 2007). The descriptive items of the WJIES categorized under both performance skills and creative development were read to the evaluator in order to provide a definitive explanation for the two traits. Six of the 12 improvised solos were chosen randomly and evaluated twice so that test-retest reliability coefficients for both traits could be calculated.



## Procedures

*Development of the Stimuli.* The 12 improvisations from the six participating performers were recorded digitally. Each performer was asked to complete two tasks. The first consisted of an improvisation of two choruses of Fast Blues in Bb (quarter note = 197 bpm) performed with a Jamey Aebersold (1981) play-along recording. Written chord changes were provided. The second improvisation task consisted of each player performing one chorus of Benny Golson's "Killer Joe" (quarter note = 115 bpm), also accompanied by an Aebersold (1979) play-along recording. One chorus of the Bb Blues accompaniment was played for each participant prior to the improvisation. The A section of the "Killer Joe" song-form also was played for each instrumentalist prior to the second improvisation task, and a lead sheet with chord changes was provided. Some participants chose to play along with the recorded accompaniments prior to recording each task in order to further familiarize themselves with the specific chord progressions. All improvisations were recorded using a Sharp IM-DR 420HS 1-bit MiniDisc recorder along with a Sony ECM-MS907 stereo microphone. The accompaniment was played on a Technics SL-PG340 compact disc player utilizing a set of Paradigm Performance Series speakers. The recordings were made in a small, acoustically treated studio office. At the conclusion of each improvisation task, the recordings were played for each participant to determine acceptability. If a performer judged an improvisation to be unacceptable, additional takes were recorded until an acceptable performance was produced.

Once the improvised solos were recorded successfully to MiniDisc, the stimuli were uploaded to computer via Apple's Garage Band program and then burned to compact disc. In an effort to provide the best possible stimuli to the adjudicators, minor balance effects were added to each recording prior to burning the disc. The improvisations were arranged on each disc in order, according to my assessment of the performer's improvisation ability, from intermediate level through sophisticated.

*Data Collection.* Before evaluating the stimuli, each judge was asked to listen to two anchor examples representing a strong and a relatively weak performance on the Bb Blues task. Each adjudicator was asked to read and adhere to specific instructions related to the anchors prior to evaluating the stimuli. The anchor recordings were provided so that each evaluator would have a reference point in order to judge each solo within the ability range of the samples. The researcher performed and recorded both examples. The first anchor example represented a relatively good solo based on a conscious attempt to address positively each of the 14 items found on the WJIES. If I used the WJIES to self-evaluate anchor recording number 1, each of the 14 items would score a 5 or higher. Likewise, the second example represented a relatively weak solo based on an attempt to address inadequately each performance category. The majority of the scores for this improvisation would fall on the lower end of the scale. At the conclusion of the recording process, the two anchor recordings also were evaluated by the researcher against the 12 stimuli in order to ensure that both examples conformed to the range of skill exemplified by the improvisations. In the



same way that you would not judge beginners with the same criteria as professionals, the raters were asked to concentrate on judging each solo within the range of skill illustrated by the two anchor recordings.

Adjudicators were given a packet containing one WJIES as well as one Instrumental Jazz Improvisation Evaluation Measure (IJIEM; May, 2003) for each recorded solo. Detailed instructions were provided by the "Directions for Evaluators" letter within each packet. The IJIEM was included for validation purposes. It consists of seven items each paired with a 7-point Likert-type scale. The seven items include technical facility, rhythm/time feel, melodic and rhythmic development, style, harmonic material, expressiveness, and creativity. Along with the evaluation forms, a compact disc containing both the recorded improvisations and the anchor recordings was provided. The adjudicators were asked to evaluate the improvisations accordingly and not to discuss the scores until all forms were returned to the researcher. The completion time for the adjudicator packet was estimated at approximately 60 minutes. The majority of evaluators returned the completed packets within 3 weeks.

## Results

The 14 items that were retained for the WJIES were selected from the perspective of a facet-rational theoretical framework first described by Butt and Fiske (1968). Consistent with a facet-rational approach to scale construction, decisions to retain, combine, modify, or reject items were made based largely on input from accomplished jazz musicians and the knowledge and expertise of the researcher. A majority (64%) of the items were adapted from descriptions of wind jazz improvisation written by knowledgeable jazz performers and educators. Other items (36%) were modified from books and magazines related to jazz improvisation and items found in the jazz improvisation research literature (Anderson, 1995; Berliner, 1994; Clarke, 1991; Coker et al., 1970; Hores, 1977; Horowitz, 1994; Hynes, 2000; May, 2003; Murph, 2005; Stamm, 2001). The final 14 items of the WJIES are found in Table 1.

Butt and Fiske's (1968) facet-rational approach calls for a grouping of items into subscales based on the researcher's a priori conceptions. Hence, I initially grouped the 14 items of the WJIES into four broad areas: (a) technique/tone quality, (b) structure/development, (c) rhythm/style, and (d) expression. The decision for this grouping was based on a combination of my assumptions related to the structure of wind jazz improvisation and a perusal of various music performance adjudication forms including both classical and jazz formats. The various evaluation methods reviewed seemed to support a four-area delineation of wind improvisation performance. I utilized the principle component analysis extraction method along with varimax rotation to factor analyze the 12 solos. Initial analyses suggested various solutions ranging from one to three factors. Seven of the 12 improvised solo evaluations produced factor solutions with three primary factors. Improvisation 8 produced a solution with a single factor. The 12 improvisations also were examined via the Kaiser-Meyer-Olkin

**Table 1.** Wind Jazz Improvisation Evaluation Scale Items

---

1.	Soloist demonstrates a knowledge of theory.
2.	Soloist plays with uncharacteristic tone quality.
3.	Soloist uses melodic motifs and/or sequences.
4.	Soloist plays with a lack of confidence.
5.	Soloist plays with appropriate time feel and/or rhythm.
6.	Soloist plays with good technical facility.
7.	Soloist expresses ideas with a lack of certainty.
8.	Soloist plays with poor intonation.
9.	Development of solo is logical.
10.	Soloist performs with emotional expression.
11.	Soloist plays with appropriate style.
12.	Soloist's performance lacks imagination and/or creativity.
13.	Solo lacks interaction and fails to dialog with accompaniment.
14.	Soloist effectively uses chromatic approach tones.

---

procedure. Results ranged from .722 to .905, a strong indication of sampling adequacy (Asmus, 1989). Based on an examination of the scree plots and rotation matrices for each of the 12 improvised solos, a two-factor structure was determined to be the best overall descriptor of collegiate wind jazz improvisation. This conclusion was reached based on the overall strength of the factor loadings for the three component initial analyses, and the distribution of items, especially those related to Factor 3. Analysis of the initial factor loadings specific to Component 3 seemed inconsistent at best. Scree plots showed a leveling off of the curve between the second and third factors, further supporting the viability of a two-factor solution. I subsequently forced a two-factor solution for each of the 12 improvised solos, all of which yielded logical item groupings. I named the two factors (a) performance skills and (b) creative development. The two factors changed place in amount of variability explained depending on the quality of the performances. For the intermediate performances, performance skills generally emerged as the first factor. When the performances were more sophisticated, creative development became Factor 1 in the analyses. Table 2 illustrates the two-factor groupings for the WJIES.

Although the a priori structure was not supported by the factor analyses, a logical reconciliation of the resulting two-factor solution is possible. The majority of the items that make up the technique/tone quality and structure/development factors within the a priori structure emerged as elements of Factor 1, performance skills. Likewise, most of the items that constitute the rhythm/style and expression factors within the a priori structure loaded on Factor 2, creative development. Factor 1, performance skills, essentially represents items related to technique and music structure. Factor 2, creative development, is made up of items specific to musical expressiveness, creativity, and style.

**Table 2.** Wind Jazz Improvisation Evaluation Scale as a Two-Factor Structure**I. Performance skills**

1. Soloist demonstrates a knowledge of theory.
3. Soloist uses melodic motifs and/or sequences.
4. Soloist plays with a lack of confidence.
5. Soloist plays with inappropriate time feel and/or rhythm.
6. Soloist plays with good technical facility.
8. Soloist plays with poor intonation.
9. Development of solo is logical.

**II. Creative development**

2. Soloist plays with uncharacteristic tone quality.
7. Soloist expresses ideas with a lack of certainty.
10. Soloist performs with emotional expression.
11. Soloist plays with appropriate style.
12. Soloist's performance lacks imagination and/or creativity.
13. Solo lacks interaction and fails to dialog with accompaniment.
14. Soloist effectively uses chromatic approach tones.

Two adjudicator panels were employed for the reliability procedures conducted within the present study. A large panel of 63 judges was composed primarily of students enrolled in two separate university jazz performance programs. The students possessed a wide variety of jazz performance, jazz education, and jazz instruction experience. A second, much smaller adjudicator panel consisted of the 4 most highly experienced jazz musicians among the original 63. A variety of strategies was used to determine the reliability of the WJIES. Cronbach's alpha ranged from .87 to .95, with 11 of 12 evaluations producing a coefficient score of .90 or better. These high reliability coefficients indicated a substantial level of internal stability in the WJIES structure.

Adjudicator consistency was calculated by an application of the Intraclass Correlation Coefficient and Kendall's Coefficient of Concordance (*W*). The intraclass correlation coefficient measured the agreement of the 63 judges within the set of 14 items. Intraclass correlation scores were comparable to the high alpha estimates, ranging from .87 to .95. The more conservative *W* was calculated both for the panel of 63 and for the smaller panel of 4. Both procedures yielded mixed results. In reference to the panel of 63, coefficients ranged from .39 to .66 for the 12 improvisations. The panel of 4 produced marginally better coefficient scores, ranging from .23 to 1.00. The high coefficient alpha values produced by the WJIES suggest that low interjudge reliability may owe to adjudicator unreliability rather than to structural flaws within the WJIES. The reliability statistics also indicated that the adjudicators more reliably evaluated the solos at the advanced and sophisticated levels. These outcomes suggest that adjudicators seem better able to agree on what constitutes skilled rather than developing improvisation performance. Complete reliability statistics are reported in Table 3.

**Table 3.** Reliability Analyses for the Instrumental Jazz Improvisation Evaluation Scale (IJIEM) and the Wind Jazz Improvisation Evaluation Scale (WJIES)

Improvisation	IJIEM			WJIES						
	$\alpha$	$W_4$	$\alpha$	IC	Avg.-IC <sub>6</sub>	SD	IC-Single	$\Lambda_1^a$	W	$W_4$
1	.85	.79**	.87	.87	.81	.18	.33	.81	.39	.71*
2	.91	.56**	.90	.90	.80	.14	.40	.84	.42*	.37**
3	.93	.63**	.91	.91	.90	.06	.42	.84	.48*	.56**
4	.93	.70**	.93	.93	.89	.08	.47	.86	.52*	.23**
5	.94	.63**	.92	.92	.90	.06	.44	.85	.53*	.82**
6	.91	.62**	.95	.95	.84	.06	.57	.88	.66*	.77**
7	.91	.56**	.90	.90	.87	.10	.40	.84	.51*	.24*
8	.93	.42**	.95	.95	.95	.03	.59	.88	.65*	.53**
9	.93	.85**	.94	.94	.92	.06	.53	.87	.64*	.86**
10	.92	.88**	.93	.93	.91	.08	.50	.87	.60*	.89**
11	.89	.71*	.93	.93	.90	.07	.50	.87	.56*	.86**
12	.91	.86**	.95	.95	.95	.02	.56	.88	.62*	1.00**

Note:  $\alpha$  = Cronbach's coefficient alpha; IC = intraclass correlation coefficient ( $n = 63$ ); Avg.-IC<sub>6</sub> = average intraclass correlation of six random extractions (average of 10 runs); SD = standard deviation; IC-Single = estimate of intraclass correlation assuming a single adjudicator;  $\Lambda_1$  = Guttman's Lambda for lower bound reliability; W = Kendall's coefficient of concordance ( $n = 63$ );  $W_4$  = Kendall's coefficient of concordance with  $n = 4$ .

<sup>a</sup> Guttman proposed several measures that estimate the lower bounds for true reliability, that is, the reliability of the hypothesized population of adjudicators. In this study, I used the first estimate, which is the simplest and most conservative.

\* $p < .05$ . \*\* $p < .01$ .

The MTMM Matrix

Construct validity was determined by the analysis of a MTMM matrix encompassing three measurement methods (the IJIEM, the WJIES, and a global rating) and two performance traits (performance skills and creative development). The performance traits emerged as a result of factor analyzing the 12 improvisation evaluations. The matrix represents an integrated multivariable platform by which information relative to convergent and discriminant validity is gathered and evaluated. With this procedure, assessments of two or more traits using two or more measurement methods are intercorrelated (Bryant, 2000). Campbell and Fiske's (1959) approach produces four different types of correlations: (a) monotrait-monomethod, (b) heterotrait-monomethod, (c) monotrait-heteromethod, and (d) heterotrait-heteromethod.

One form of construct validity found in MTMM designs specifically relates to convergent validity. Convergent validity is the degree to which multiple measures of the same construct agree, or converge (Campbell & Fiske, 1959). Valid measures of the same underlying concept should correlate highly.

Campbell and Fiske (1959) proposed several criteria for the evaluation of validity within the MTMM matrix. The first of these criteria deals with convergent validity.

**Table 4.** Multitrait-Multimethod Matrix

	Trait	A1	B1	A2	B2	A3	B3
Method 1: IJIEM	A1	<b>.76<sup>a</sup></b>					
	B1	<b>.98**</b>	<b>.82<sup>a</sup></b>				
Method 2: WJIES	A2	<b>.96**</b>	<b>.98**</b>	<b>.89<sup>a</sup></b>			
	B2	<b>.98**</b>	<b>.99**</b>	<b>.97**</b>	<b>.87<sup>a</sup></b>		
Method 3: Global rating	A3	<b>.78**</b>	<b>.81**</b>	<b>.76**</b>	<b>.73**</b>	<b>.50<sup>b</sup></b>	
	B3	<b>.79**</b>	<b>.84**</b>	<b>.83**</b>	<b>.79**</b>	<b>.80**</b>	<b>.76<sup>b</sup></b>

Note: Traits: A = performance skills subscale, B = creative development subscale. The monotrait-monomethod diagonal is in boldface.

<sup>a</sup> Averaged coefficient alpha for the 12 evaluations.

<sup>b</sup> Test-retest correlations.

\*\*p < .01.

Convergent validity exists when the correlations among multiple methods of measuring the same construct (monotrait-heteromethod coefficients) are “significantly different from zero and sufficiently large” (p. 82). These values correspond to the correlations between A1-A2, A2-A3, A1-A3, B1-B2, B2-B3, and B1-B3 found in Table 4. The corresponding coefficients of .96, .76, .78, .99, .79, and .84, respectively, confirm this criterion. Therefore, convergent validity of the WJIES is confirmed.

Content validity was determined via two methods: (a) a methodical development of the initial item pool and (b) a later validation of the structure of the items by factor analysis. By incorporating the expertise of a wide variety of authorities in the development of the initial item pool, a comprehensive description of wind jazz improvisation was produced. The resulting 14 items of the WJIES were factor analyzed, resulting in two distinct factors that subsumed the four-area a priori structure.

Criterion-related validity was examined by comparing total scores obtained from the WJIES and the IJIEM evaluations. Pearson’s *r* was used to analyze relationships between each pair of summated scores. Results were mixed. Six of the 12 comparisons produced low to moderate correlations ranging from .24 to .89, while the remaining six comparisons yielded high correlations ranging from .91 to 1.00. While the highest three correlations (.99, .99, 1.00) occurred between improvisations at the advanced and sophisticated levels, there does not seem to be a discernable pattern to these results. Perhaps, variability in adjudicator experience is responsible for the inconsistent outcomes. Although the correlation outcomes were mixed, the 1.00 found for arguably the best of the 12 performances is encouraging.

**Discussion**

Jazz education in America has progressed steadily since its beginnings in the mid-1940s (Abeles, Hoffer, & Klotman, 1995). Horowitz (1994) successfully predicted that

“as this growth continues, there will be a concomitant increase in the need for objective methods of evaluation” (p. 67). A multimillion-dollar industry has emerged that supplies a large quantity of pedagogical materials to jazz educators as well as to students of all ages. Jazz theory books, play-along materials, and various audiovisual methods are widely accessible to aspiring jazz musicians. Unfortunately, the ready availability of pedagogical materials has not led to the development of a wealth of valid and reliable methods of assessing improvisation. In spite of this, evaluations of jazz improvisations must continue to take place (Horowitz, 1994). Auditions, contests, festivals, studio lessons, and private practice sessions illustrate the various situations in which reliable and valid assessment procedures would assist student progress.

Although the present study is similar in many ways to the investigations conducted by Horowitz (1994), May (2003), and McPherson (1995), it is a departure from the facet-factorial approach to scale development. Instead, I employed a facet-rational strategy to develop the WJIES. The results of this study demonstrate that the facet-rational approach to rating scale construction can be used to develop a wind jazz improvisation measure that exhibits acceptable levels of both reliability and validity. Music educators who possess extensive instructional and performance experience, aided by a systematic method, should be able to develop effective performance measures. Jazz improvisation performance is in a constant state of transformation. Therefore, evaluation of such an evolving construction must be flexible and adaptive. The evaluator of an *avant-garde* or *free* wind jazz improvisation might need to consider a set of variables unique to those subcategories of jazz. This might be true especially when considering improvisation that departs from tradition harmonic parameters. The facet-rational approach to scale construction may provide jazz educators with the means to introduce a degree of adaptability and flexibility that might be missing in a facet-factorial based approach to scale development. Horowitz (1994) stated the following:

Many rating scales at festivals are developed through the “rational” means described by Butt and Fiske (1968). That is, they are constructed according to the opinions of the developer. This may not result in objective and valid measures. (p. 71)

I am inclined to agree. But I would argue that years of specialized music expertise should not be discounted. Utilized within a methodical facet-rational approach to scale development, the extensive knowledge of expert teachers could be used to produce objective performance measures.

One of the primary outcomes of any scale development project is an eventual description of the construct. In this case, the 14-item WJIES along with previously constructed jazz improvisation scales (Horowitz, 1994; May, 2003; McPherson, 1995; Pfenninger, 1990) expand the collective understanding of the jazz improvisation process. In contrast to May (2003), who found that instrumental jazz improvisation emerged as a single construct, the results of this study suggest a



two-factor structure. The present findings also represent a departure from the work of Horowitz (1994) and Madura (1996), whereas jazz improvisation performance was defined as a three-factor construct. These varied results could be viewed as confounding jazz improvisation understanding. May (2003) suggested that mixed results might be due to differences in performance mediums or, rather, the selection of differing criteria at the start of the investigations. I believe the combined findings suggest that jazz improvisation is an extremely complex construct that might be understood better within specific contexts. Continued research focused on revealing what core elements constitute jazz improvisation in a variety of contexts is needed.

Further descriptions of jazz improvisation also may help music educators address pedagogical issues related to improvisation in general. Although improvisation is linked inherently to jazz, it also is associated with many different genres of music performance (Bitz, 1998; Tomassetti, 2003). Elementary music is one obvious example. Children routinely are taught to improvise melodies both vocally and instrumentally. Although the WJIES was developed with more mature performers in mind, many of the performance skills described by the scale might be applied to a variety of skill levels including elementary-aged students. Through transfer and modification, it may be possible for music educators teaching a variety of levels and genres to utilize items found within the WJIES successfully for the purpose of evaluating improvisation. The scale also may be useful to teachers developing instructional strategies designed to teach improvisation. Additional investigations are needed in order to confirm or refute these suggestions.

A critical finding of this study was the emergence of two factors central to wind jazz improvisation—performance skills and creative development—and the factors' relationship to skill level. May (2003) concluded that if the subskills that contribute to instrumental jazz improvisation achievement are truly interdependent, they should be developed "contiguously." However, the results of the present study support a sequential format for teaching jazz improvisation instruction implicated by the close relation of the performance skills factor and developing improvisation. The items associated with the creative development factor were related more closely to improvisation at the sophisticated level. Consequently, it is reasonable to suggest that in order to progress from a novice improviser to a sophisticated or expert improviser, the elements that make up the performance skills factor must be mastered first. This is consistent with the conclusions reached by Antonelli (1997), Bash (1983), Burnsed (1978), and Meadows (1991). The elements of the performance skills factor include jazz theory, melodic motifs and/or sequences, confidence, time feel, technique, intonation, and solo development. Once a progressing improviser is relatively proficient at the skills related to this factor, those skills can be transcended and attention can be focused on the elements of creative development, which include fluidity, expression, imagination and/or creativity, and so forth. Works completed by Berliner (1994) and Corpolongo (1995a, 1995b, 1996a, 1996b, 1997) outline methodology for the development of the skills associated with creative development.

There is one issue associated with the hypothesized “sequential format” for teaching improvisation that must be considered, however. As stated earlier, the improvisation stimuli were placed on the adjudicator compact disc according to my perceptions of the improvisation ability of the performers. Instrumentalists at the intermediate level performed solos 1 through 4. Advanced-level performers recorded solos 5 through 8, and instrumentalists with sophisticated skill performed solos 9 through 12. This does not mean inevitably that the stimuli produced strictly followed this delineation. But it does increase the possibility of an order effect. Avoiding the possibility of an order effect would be paramount to investigating the sequential format hypothesis.

Improvisation should not be deferred until sufficient performance skills are realized, however. Students of all ages should be encouraged to experiment with improvisation early in their development. These skills must be learned by engaging actively in the improvisation process. What the findings of this study do imply, however, is that the level of sophistication may be related to performance skill. As performance skill increases, the potential to create more creative improvised solos also may increase. Therefore, any competent musician can learn to improvise. Improvisation is not a cryptic talent bequeathed only to a chosen few. Advanced improvisation skill can be learned if students are willing to develop the prerequisite skills.

In addition, jazz educators might foster advanced improvisation skills successfully by developing curricula that focus on performance skills early in the instructional process. One of the first elements that must be cultivated in young students is confidence. Improvisation is essentially a leap of faith in which musicians rely solely on their referent knowledge and expression to create spontaneously. This requires sufficient confidence. If young students are unsuccessful at developing this confidence early, it may become increasingly difficult to foster, especially relative to female instrumentalists (Webster, 1977; Wehr-Flowers, 2007).

There is one point of contention related to the two-factor solution that deserves mention. Item number 2, *Soloist plays with uncharacteristic tone quality*, consistently was grouped with Factor 2, creative development. Some would argue that tone quality should be appropriately grouped with Factor 1, performance skills. The factor analysis did not support this grouping. Maybe the judges somehow viewed tone quality in a more subjective manner as a result of the jazz context. This outcome is plausible especially when considering the possible variability of saxophone tone. For example, the difference between Lester Young’s tenor saxophone tone and John Coltrane’s tenor saxophone tone is quite significant. But to place value on that difference invites a certain amount of subjectivity. By its very nature, jazz allows for a wide spectrum of “acceptable” tone qualities. So in essence, in the context of jazz, tone might indeed function as a “creative” variable. Further investigation of this specific query is needed.

Furthermore, the results related to judge reliability underscore several issues related to adjudicator training. The adjudicators in the present study were largely

student performers with little or no experience evaluating jazz improvisation. The variability of the adjudication scores suggests that mere participation in jazz performance ensembles does not prepare musicians sufficiently to adjudicate improvisation reliably. Although anchor recordings were provided for each judge, the resulting scores varied. This outcome suggests a need for more extensive training for jazz improvisation adjudicators, especially in the case of inexperienced judges. The findings of this study also indicate a need for training in the case of experienced jazz musicians. Given that an acceptable jazz improvisation can be manifested in a variety of forms, determining what constitutes a good performance may be problematic even for veteran judges (Pfenninger, 1990). Adjudicator training has become a staple of solo and ensemble competitions across the country. Some form of structured adjudicator training prior to jazz festivals, jazz tryouts, and so forth seems reasonable.

As previously stated, adjudicator consistency was a concern in the present study. Based on the reliability outcomes of the WJIES, researchers should explore reliability among various panels of judges. How reliable is the WJIES when used by a larger panel of experienced adjudicators? Is reliability adversely affected when used by judges possessing extensive instrumental music education credentials but who lack specific jazz education experience? How does adjudicator training affect reliability?

Although the WJIES was developed with a specific age group and ability level in mind, the WJIES might be effective in evaluating performances from a wider age range and variety of improvisation skill levels. By examining the reliability of the scale across varying ability levels, the efficacy of using the WJIES to evaluate developing improvisers may be confirmed or refuted.

Future research should be undertaken to examine findings in relation to previously conducted research in order to advance understanding of the jazz improvisation construct and to determine whether WJIES items support or contradict theoretical explanations for the improvisation process. By doing so, a more productive approach to jazz instruction, especially as it relates to jazz improvisation, might be developed.

### **Author's Note**

The article is based on the author's doctoral dissertation, "Development and Validation of a Rating Scale for Wind Jazz Improvisation Performance," completed at the University of Missouri–Columbia in 2007.

### **Declaration of Conflicting Interests**

The author declared no conflicts of interest with respect to the authorship and/or publication of this article.

### **Funding**

The author received no financial support for the research and/or authorship of this article.

## References

- Abeles, H. F. (1973). Development and validation of a clarinet performance adjudication rating scale. *Journal of Research in Music Education*, 21, 246–255.
- Abeles, H. F., Hoffer, C. R., & Klotman, R. H. (1995). *Foundations of music education* (2nd ed.). New York: Schirmer Books.
- Aebersold, J. (1979). Killer Joe. On *Benny Golson* [CD]. New Albany, IN: Jamey Aebersold Jazz.
- Aebersold, J. (1981). Fast blues in Bb. On *Nothing but blues* [CD]. New Albany, IN: Jamey Aebersold Jazz.
- Aitken, A. E. (1975). A self-instructional audio-imitation method designed to teach trumpet students jazz improvisation in the major mode. *Dissertation Abstracts International*, 37(01), 18A. (University Microfilms No. 7615018)
- Anderson, R. (1995). Playing and teaching improvisation. *Jazz Educators Journal*, 27, 33–35.
- Antonelli, J. K. (1997). Improvisation for beginners. *The Instrumentalist*, 51, 44–49.
- 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*, 101, 1–29.
- Bash, L. (1983). The effectiveness of three instructional methods on the acquisition of jazz improvisation skills. *Dissertation Abstracts International*, 44(07), 2079A.
- Berard, M. (1998). Production and evaluation of a self-instructional method for teaching jazz guitar. *Masters Abstracts International*, 37(06), 1605.
- 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. *Dissertation Abstracts International*, 49(05), 1086A.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137–141.
- Berliner, P. F. (1994). Thinking in jazz: Composing in the moment. *Jazz Educators Journal*, 26, 28–34.
- Bitz, M. (1998). Teaching improvisation outside of jazz settings. *Music Educators Journal*, 84(4), 21–26.
- Bryant, F. B. (2000). Assessing the validity of measurement. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 99–146). Washington, DC: American Psychological Association.
- Burnsed, C. V. (1978). The development and evaluation of an introductory jazz improvisation sequence for intermediate band students. *Dissertation Abstracts International*, 39(08), 4578A.
- Butt, D. C., & Fiske, D. W. (1968). A comparison of strategies in developing scales for dominance. *Psychological Bulletin*, 70, 505–519.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105.
- Clarke, E. F. (1991). Generative processes in music. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 1–26). Oxford, UK: Clarendon Press.
- Coker, J. (1964). *Improvising jazz*. Englewood Cliffs, NJ: Prentice Hall.
- Coker, J., Casale, J., Campbell, G., & Greene, J. (1970). *Patterns for jazz* (3rd ed.). Miami, FL: Warner Bros.

- Corpolongo, R. (1995a). Moving beyond basic improvisation. *The Instrumentalist*, 49, 46–50.
- Corpolongo, R. (1995b). Teaching improvisation with blues and bebop. *The Instrumentalist*, 50, 58–61.
- Corpolongo, R. (1996a). Advanced improvisation techniques. *The Instrumentalist*, 50, 57–58.
- Corpolongo, R. (1996b). Teaching advanced improvisation. *The Instrumentalist*, 51, 54–55.
- Corpolongo, R. (1997). Turnaround patterns in jazz improvisation. *The Instrumentalist*, 51, 50–57.
- Damron, B. L. (1973). The development and evaluation of a self-instructional sequence jazz improvisation. *Dissertation Abstracts International*, 35(02), 1141A. (University Microfilms No. 7418028)
- 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 music performance. *Dissertation Abstracts International*, 41(04), 1462.
- Gorsuch, R. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Greennagel, D. J. (1994). A study of selected predictors of jazz vocal improvisation skills. *Dissertation Abstracts International*, 55(08), 2201.
- Gridley, M. C. (1991). *Jazz styles: History & analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Hores, R. G. (1977). A comparative study of visual- and aural-orientated approaches to jazz improvisation with implications for instruction. *Dissertation Abstracts International*, 39(04), 2122A. (University Microfilms No. 7818637)
- Horowitz, R. A. (1994). *The development of a rating scale for jazz guitar improvisation performance*. Thesis (Ed. D.), Columbia University, Teachers College.
- Hynes, T. (2000). Melody and melodicism in the teaching of jazz improvisation. *Jazz Educators Journal*, 32, 46–49.
- Johnson-Laird, P. N. (1991). Jazz improvisation: A theory at the computational level. In P. Howell, R. West, & D. Cross (Eds.), *Representing musical structure* (pp. 291–325). London: Academic Press.
- Jones, H. (1986). An application of the facet-factorial approach to scale construction in the development of a rating scale for high school vocal solo performance. *Dissertation Abstracts International*, 47(04), 1230.
- Kenny, B. J., & Gellrich, M. (2002). Improvisation. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 117–134). New York: Oxford University Press.
- Kynaston, T. P., & Ricci, R. J. (1978). *Jazz improvisation*. Englewood Cliffs, NJ: Prentice Hall.
- 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.
- May, L. F. (2003). Factors and abilities influencing achievement in instrumental jazz improvisation. *Journal of Research in Music Education*, 51, 245–258.
- McPherson, G. (1995). Evaluating improvisational ability of high school instrumentalists. *Bulletin of the Council for Research in Music Education*, 119, 11–20.
- Meadows, E. S. (1991). Improvising jazz. *Music Educators Journal*, 78(4), 41–45.
- MENC—The National Association for Music Education. (2001). Gender trends among MENC music educators. *Teaching Music*, 8(6), 52–53.

- Murph, J. (2005, February). Control room. *Downbeat*, pp. 42–44.
- Pazitka-Munroe, W. L. (2002). The construction and validation of an audition instrument to measure the vocal performance of college singers auditioning for choral ensembles. *Dissertation Abstracts International*, 63(08), 2821.
- Pfenninger, R. C. (1990). The development and validation of three rating scales for the objective measurement of jazz improvisation achievement. *Dissertation Abstracts International*, 51(08), 2674A.
- Pressing, J. (1987). The micro- and macrostructural design of improvised music. *Music Perception*, 5, 133–172.
- 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. (2007). *Development and validation of a rating scale for wind jazz improvisation performance*. Columbia: University of Missouri–Columbia. Retrieved from <http://etd.missouri.edu/Summer2007/Dissertation/SmithD-073007-D7758/>
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.
- Stamm, M. (2001). Improvisation. *Jazz Educators Journal*, 34, 77–78.
- Tomassetti, B. (2003). Beginning blues improvisation pedagogy for the non-jazz specialist music educator. *Music Educators Journal*, 89(3), 17–22.
- Webster, P. R. (1977). *A factor of intellect approach to creative thinking in music*. Rochester, NY: University of Rochester.
- Wehr-Flowers, E. (2007). Differences between males and females in confidence, anxiety, and attitude towards learning jazz improvisation. *Journal of Research in Music Education*, 54, 337–349.
- Whyton, T. (2006). Birth of the school: Discursive methodologies in jazz education. *Music Education Research*, 8(1), 65–81.
- Zdzinski, S. F. (2002). Development and validation of a string performance rating scale. *Journal of Research in Music Education*, 50, 245–255.

## Bio

**Derek T. Smith** is assistant professor of music at Lincoln University of Missouri. His research interests include jazz improvisation and instrumental pedagogy.

Submitted August 29, 2008; accepted June 3, 2009.