

**The Effect of Instruction on Sixth Grade Band Students' Abilities to Self-Rate
Etude Performance**

Nathan B. Kruse
Michigan State University
School of Music
116 Music Building
East Lansing, MI 48824

krusenat@msu.edu

The Effect of Instruction on Sixth Grade Band Students' Abilities to Self-Rate Etude Performance

Abstract

The purpose of this study was to examine the effects of rating scale instruction on self-evaluation accuracy among student musicians. Sixth grade band students ($N = 36$) from a Midwestern state performed and recorded an original etude and then critically evaluated their own rhythmic accuracy using a researcher-constructed rating scale. Control and treatment groups were then created using an expert panel's evaluation of the recorded etudes.

One week later, the treatment group received instruction in how to use the rating scale and then rated their original recorded performances again. The control group received no training but also rated their etudes a second time. Inter-judge reliability, control and treatment group correlations, means, standard deviations, and standard errors of measurement were calculated using Pearson product-moment correlations.

Results indicated that rating scale instruction was more effective than no rating scale instruction in helping students improve self-rating accuracy. While the control group tended to rate themselves the same during the second listening, the treatment group tended to rate themselves more critically. Additionally, the treatment group's tendency to rate closer to the experts' ratings suggests that rating scale instruction may not only benefit students' self-evaluation accuracy, but may also be a practice strategy toward improving students' independent musicianship.

The Effect of Instruction on Sixth Grade Band Students' Abilities to Self-Rate Etude Performance

Objectivity in measuring music performance has been, to a large degree, ancillary to subjective musical performance (Gordon, 2002). Teachers in the arts have traditionally employed subjective measurements to assess student accomplishments; yet, the facets of objective measurements based on approved standards do indeed lend themselves to use in the music classroom. High concentration levels of objectivity exist when there is agreement among teachers on a given musical performance, and properly constructed rating scales may aid in this endeavor (Gordon, 2002).

Because rating scales may be used to assess a wide variety of developmental results (Linn and Miller, 2005), their strength as a measuring tool lies in the ability to evaluate one task or characteristic at any one time, and can include one or several assessors. Not only do rating scales communicate to students what constitutes an exceptional performance, they also serve as sound teaching devices for improving instruction. To this extent, it may also be useful for students to rate themselves following a task completion and compare their scores with the teacher's (Linn and Miller, 2005). When constructed properly, rating scales can serve to assess student achievement and the quality of instruction.

The purpose of this study was to investigate how well students self-evaluate, and to determine whether self-rating abilities can be developed in young instrumentalists.

While music assessment is a vital part in any instrumental instructional process, there are inherent problems in judging musical performance.

Fiske (1978) noted that judge consistency tended to be in agreement only 25% of the time. In order to correct this, he suggested either creating a panel of judges for the assessment objectives or implementing training for the judges. Fiske (1978) also noted, however, that a panel of judges is not always possible or cost-efficient, and that training alone cannot guarantee evaluation consistency.

While surveys have historically been used to hone educational directives (Bergee, 1987), researchers have studied ways in which to increase the reliability and validity components within systematic measures of musical performance (Zdzinski, 1991). Watkins and Farnum (1954) created the Watkins-Farnum Performance Scale for band instruments and, in 1969, the Farnum String Scale was adapted. While both the band and string measures proved to be highly reliable (Stivers, 1972), there were still validity issues, in part, because there were no evaluative components that measured intonation, tone quality, or interpretation. Abeles (1973) also endeavored to improve the validity of adjudication performance scales by recommending that judges be provided with systematic rating procedures to diffuse subjectivity during evaluation. Judge selection has proved to be a notable criterion (Fiske, 1975) and was further delineated by Fiske into diagnostic adjudication venues and selection-rejection situations, which both required a suitable balance between judge expertise and a properly constructed criterion measure.

Further research resulted in the construction of specific criteria to measure learning objectives. Researchers (Abeles, 1973; Fiske, 1975; Gutsch, 1965; Kidd, 1975) created measures for clarinet performance, trumpet performance, rhythmic sight-reading

accuracy, and trombone performance skill, respectively, while Saunders and Holohan (1997) developed continuous and additive rating scales in their endeavor to construct criteria-specific evaluations.

As informative feedback and opportunities for correcting errors are two notable components that increase musicianship (Ericsson, 1997), using assessment tools such as rating scales may aid in the development of musicianship among young instrumentalists. Jorgensen (1995) depicted the process of music learning as “self-teaching,” and developed a three-phase system that included planning, practice, and evaluation. While the planning and practice phases included increasingly focused levels of musical preparation, the evaluation process involved assessing the musical performance, the learning process behind the performance, and the method of self-teaching (Jorgensen, 1995). Hewitt (2001) recommended that if independent musicianship is to be achieved, students must be able to effectively evaluate their own performances through a variety of practice strategies. In this manner, students would be able to contribute more decision-making skills during ensemble rehearsals and would rely less on coaching from the band director.

Although there is limited research in the domain of self-evaluation and its effect on musical performance (Hewitt, 2001), a number of studies do provide some information examining the relationship between self-evaluation and students’ performances. In two studies, after receiving training in self-evaluation methodology, elementary students improved their ability during performance (Davis, 1981; Sparks, 1990). Although middle school students produced inconsistent ratings during self-evaluation processes, Aitchison (1995) discovered that, while their evaluation accuracy

increased over time, their performance ability did not increase as much as the students who received teacher feedback. These findings were consistent with prior research (Colwell, 1995; Rosenthal, 1985) that respectively reported improved self-evaluation accuracy among students who received both peer evaluation and teacher feedback. Furthermore, students may be encouraged to accept more responsibility for their own learning through continued self-evaluation practices (Bergee and Roberts, 2002).

Several studies (Aitchison, 1995; Davis, 1981; Hewitt, 2001; Sparks, 1990;) reflected that self-evaluation might have an effect on students' attitudes toward themselves. While Davis (1981) found positive attitudes among beginning band students and their perceptions of self-evaluation, Sparks (1990) also reported positive attitudes toward the band director and the music classroom in general. Aitchison (1995) concluded that self-evaluation encouraged intrinsic musical interests and in the perception of musical performance.

An essential component of self-evaluation has been described as judging self-monitored information against a given standard or goal (Davidson & Scripp, 1992; Linn and Miller, 2005; Slavin, 1991; Zimmerman, 1998). In this way, musicians may be able to rely upon internal and external models to use as a comparison. Because internally generated models have been identified as being ancillary to external models, live or recorded performances have been found to be more accurate and reliable than non-modeled performances in increasing students' self-evaluation accuracy (Bundy, 1987; Kepner, 1986). Although Kepner (1986) did not indicate specific inaccuracies, such as pitch or rhythm in error detection, he concluded that students were better able to detect errors they made using audiotapes versus hearing live performances. Bundy (1987) found

that live performances served students better in their ability to accurately detect errors associated with pitch than rhythm. These findings suggest that, although students are able to detect pitch errors during their own live performances, audiotapes may be superior in identifying other types of musical errors during performance.

Hewitt (2001) focused on the self-evaluation tendencies of junior high students as they related to the inclusion or exclusion of modeled recordings. One of the leading findings of the study was that students might make inaccurate assessments of their own performance if they have not compared it with a model. The author of the study noted that performance areas such as rhythmic accuracy, melodic accuracy, tone, and interpretation could be improved when self-evaluation was coupled with an external model. Self-evaluation by itself with no external model was considerably less effective in improving student performance.

Using cognitive research models, self-regulated learning (SRL) is shown to exist in operant and social-cognitive domains, which includes the self-regulatory processes musicians experience (McPherson and Zimmerman, 2002). Zimmerman (2000) developed an SRL model that incorporated forethought, performance, and self-reflection. The latter included components of self-judgement and self-satisfaction. Comparing self-evaluation processes against established criteria appeared to be a fundamentally sound technique, as fixed standards highlighted what students had actually learned (Covington & Roberts, 1994). In some studies not involving music, some educational researchers indicated that the cognitive immaturity of children might have prevented them from self-evaluating accurately (Eshel & Klein, 1981; Nicholls, 1984; Nicholls & Miller, 1983), although this was not always the case in terms of musical self-evaluation. For example,

although college brass students and college conductors generally rated themselves higher in self-evaluations as compared with expert evaluators (Bergee, 1993; Byo, 1990), Aitchison's research led him to believe that middle school students were able to increase their ability to self-evaluate over an extended period of time, especially when teacher feedback was involved (Aitchison, 1995). Hewitt (2001) reported that combining aural models with self-evaluation techniques appeared to be more effective than isolating the aforementioned procedures in increasing overall music performance (except articulation, tempo, and intonation).

Drawing on David Elliott's sentiment regarding music educators' obligation to teach students how to "continue developing their musicianship in the future" (Elliott, 1995, p. 261), Hewitt (2002) created a study that measured junior high students' abilities to self-evaluate musical performance. The results indicated that self-evaluation abilities increase over time, except in the area of intonation, for which accuracy in self-evaluation may actually decrease with time. Except for the area of technique/articulation, students positively rated their performance and, as supported by previous findings of conductors and college brass students (Bergee, 1993; Byo, 1990), tended to rate themselves higher than the experts.

The aforementioned research poses several implications for improving classroom instruction and raising students' self-awareness regarding specific musical tasks. Of particular interest to this study is that of student comprehension in using a rating scale, its language use and design, and how self-rating accuracy may be improved. As previously mentioned, rating scales may be used to assess a wide variety of developmental results and can include one or more assessors. To this extent, accurate assessment may only

occur when assessors understand and correctly interpret the language used within rating scales and apply it to a musical performance.

The specific problems of this study were 1) to identify how sixth grade instrumental music students rate themselves in relation to a panel of expert judges, and 2) to determine the level of effectiveness that instruction may have in developing self-rating skills among young instrumentalists.

METHOD

Subjects

Subjects in this study were sixth grade band students from an intermediate school in a Midwestern state. Thirty-six students (fourteen boys and twenty-two girls) participated in the study and consisted of woodwind ($n=13$), brass ($n=19$), and percussion ($n=4$) players. Students in the instrumental music program had been playing their instruments for eight months and used a band method that initially addressed the tonal and rhythmic aspects of music, and placed special emphasis on improvisation. The introductory teaching stages of this band method did not include notation, and at the time of this study, the students had been reading notation for five months.

Preparation of Materials

The researcher composed an original, eight-measure etude, which is shown in Figure 1. It was written in a familiar key to the students and included a variety of rhythmic patterns that were common to the sample population of students. Several familiar tonal patterns the students typically performed during class instruction were also used as melodic material for the etude. Two music experts, one of whom was the school

band director, examined the etude and concluded it was appropriate for use with this group of students.

Figure 1
Performance Etude

Yellow and Blue Make Green

Kruse

B♭ Clarinet

A one-dimension continuous rating scale containing five criteria was constructed for use in measuring the rhythmic accuracy of the performances. While students' rhythmic aptitude was neither measured nor deemed a determining factor in the present investigation, rhythm was selected as the solitary dimension for the study because of its objective attributes. The investigator-constructed scale was designed for use in both the pre- and post-training segments of this study:

- 5 = performs all notated rhythms accurately and in tempo
- 4 = performs all notated rhythms accurately with slight tempo alterations
- 3 = performs most rhythms accurately
- 2 = performs a few notated rhythms accurately but most of them inaccurately
- 1 = plays with no rhythmic accuracy

A portable cassette recorder with advanced recording capabilities was used to record student performances. One cassette tape was employed for recording the woodwind students and a second cassette was used for recording the brass and percussion students.

Design and Procedures

The project was designed to include six school visits over a two-week time period. The school's predetermined class schedule provided a natural division of two student sections: woodwinds and brass/percussion. During the first week, participating students individually entered a large ensemble room and were asked to study the etude silently for sixty seconds while the researcher provided a steady beat. Before being recorded, students were instructed to sight-read and perform the etude once in its entirety without stopping. Then, listening to their own taped performances, students critically evaluated their own rhythmic performance using the rating scale. Three expert judges subsequently rated the same performances from the cassette tapes, and, using these ratings, two student groups (control, $n = 18$, and treatment, $n = 18$) were then created by the researcher, controlling for balanced instrumentation and performance achievement. During the second week, the treatment group received instruction in how to use the rating scale. Instruction consisted of defining and discussing the language used in the rating scale and hearing examples of correct and incorrect examples of the etude in relation to the rating scale. Students in the treatment group were then asked to rate for a second time their original recorded performance from the previous week. The control group received no instruction but also rated their original recorded etude performance again. Inter-judge reliability, control and treatment group correlations, means, theoretical and observed standard deviations, and standard errors of measurement were then calculated from the collected data.

RESULTS

Using a Pearson product-moment correlation, acceptable inter-judge reliabilities were found between the three expert judges. The inter-judge reliability was .70 between Judge 1 and Judge 2, and .75 between Judge 1 and Judge 3. A higher inter-judge reliability of .81 was found between Judge 2 and Judge 3.

Table 1 provides the means, standard deviations, and standard errors of measurement found in the students' first listening, second listening, and the expert judges' reliability. In each case, the observed mean (OM = 2.80 and 2.69, respectively) was lower than the theoretical mean ($M = 3$). After the first listening, the treatment group rated themselves similarly in rhythmic accuracy to the control group. After the second listening, the treatment group's mean score ($M = 2.61$) fell in relationship to the control group ($M = 2.78$), whose mean score actually increased following the second listening.

The composite judge scores were added then divided by three in order to provide comparable results. The expert judges rated the treatment group ($M = 2.50$) slightly higher in rhythmic accuracy compared to the control group ($M = 2.43$), and the experts' observed mean was less than the theoretical mean ($M = 3$). While the control group's mean, standard deviation, and standard error of measurement between Listening 1 and Listening 2 were identical, the results for the treatment group reflected a change in self-evaluation.

Table 1

Student Listening 1 and 2/Composite Judge Results: Means, Standard Deviations, and Standard Errors of Measurement

	<i>TM</i>	<i>OM</i>	<i>TSD</i>	<i>OSD</i>	<i>SEM</i>
Student Listening 1, All Subjects	3	2.80	.67	.82	.14
Student Listening 1, Control	3	2.78	.67	.88	.21
Student Listening 1, Treatment	3	2.83	.67	.79	.19
Student Listening 2, All Subjects	3	2.69	.67	.82	.14
Student Listening 2, Control	3	2.78	.67	.88	.21
Student Listening 2, Treatment	3	2.61	.67	.78	.18
Composite Judge, All Subjects	3	2.46	.67	.87	.13
Composite Judge, Control	3	2.43	.67	.74	.17
Composite Judge, Treatment	3	2.50	.67	.79	.19

Table 2 provides correlations between the ratings of the control and treatment groups and the ratings of the expert judges from the first and second listening sessions of the etude. For the control group, the relationship decreased from .73 to .52. In contrast, the treatment group increased the relationship of their ratings to those of the experts from an exceptionally low .33 to a moderately acceptable .63. A T-test was used to look for significant differences between groups in their own pre- and post-treatment ratings, and no differences were found.

Table 2

Correlations between Expert Panel and Control/Treatment Groups

	<i>Expert Judges and Student Listening 1</i>	<i>Expert Judges and Student Listening 2</i>
Control	<i>r = .73</i>	<i>r = .52</i>
Treatment	<i>r = .33</i>	<i>r = .63</i>

DISCUSSION

Instruction in how to use a rating scale was found to be more effective than no instruction in how to use a rating scale in helping students critically self-evaluate their etude performance. Each of the groups (control and treatment) rated themselves lower than the experts' judgments. However, while the control group's assessment of their performances did not change as a result of the second listening, the treatment group tended to rate themselves more critically on the second listening. The tendency of the treatment group to rate closer to the experts' ratings suggests that students benefit from rating scale instruction in the music classroom, as this instruction may increase the accuracy of self-evaluation practices. This self-teaching approach (Hewitt, 2001; Jorgensen, 1995) is consistent with existing research (Davis, 1981; Hewitt, 2002; Sparks, 1990) that addresses improved musical ability in the classroom through increasingly accurate self-evaluation skills. When paired with teacher feedback, this self-reflective process can yield improved effectiveness (Ericsson, 1997).

The differences in the observed, theoretical, and expert means in this study suggest the etude may have been too difficult for the students. Also, the broad concept of rhythmic accuracy in relationship to tempo fluctuations as performed by a majority of the students may have contributed to the varied, although acceptable, reliability among the expert raters. While students in the treatment group ultimately rated themselves more accurately following instruction in how to use the rating scale, neither the treatment group nor the control group excelled in their etude performance.

Students in the treatment group responded positively to instruction and remained focused and engaged throughout the second listening. This may have been a result from

instruction in using the assessment tool and, from the purposeful attention of the researcher, a general sense of individual ownership and investment in the self-evaluation process. Several studies (Aitchison, 1995; Davis, 1981; Hewitt, 2001; Sparks, 1990) addressed students' positive attitudes toward and higher ratings in self-evaluation procedures. While Davis (1981) presented affirmative attitudes among beginning band students and their perceptions of self-evaluation, Sparks (1990) also reported positive attitudes toward the band director and the overall music classroom. In general, students in the present study tended to rate themselves higher than the expert panel of judges, which mirrors the patterns found in other studies (Bergee, 1993; Byo, 1990; Hewitt 2002).

The deficit in ability to self-evaluate appears to be persistent, and improving this construct will require studying the interplay between the inherent and environmental influences students face (Bergee and Roberts, 2002). While the tendency of the treatment group to rate closer to the experts' ratings suggests that students benefit from rating scale instruction, student indifference and a reduction of purpose and meaning in the self-evaluation process by the control group was noted. Another reason for this anomaly may have been the time of day the subjects were tested and the subsequent levels of the participants' focus. While students in the woodwind class participated in the study following lunch and recess activities, brass and percussion students participated at the end of the school day. Unavoidable distractions associated with these timeframes within the school day may have affected the level of concentration in both groups.

While some studies have shown that instruction alone did not increase evaluation consistency (Colwell, 1995; Ericsson, 1997; Fiske, 1978; Rosenthal, 1985), Hewitt (2002) indicated that self-evaluation abilities can increase over time and recommended

that clear instructions be provided each time a student completes a self-evaluation task, especially if younger players are involved. This is a compelling notion, as the ability to self-evaluate is a National Standard and is a critical skill for both individual musicianship and for understanding how one meets the expectations of a performance ensemble or instructor. Furthermore, junior high band directors should continually instruct students in self-evaluation measures so that systematic processes are instilled and maintained. Implementation of such a construct is recommended in the early stages of developmental learning (Brown, 1999).

Although the current study illustrates that instruction in how to use a rating scale was more effective than no instruction in how to use a rating scale in helping students critically self-evaluate their etude performance, students were unable to effectively and consistently self-evaluate their individual music performances. Using a larger sample size and measuring students' rhythmic aptitude may yield additional interpretations in future studies. Another consideration of particular interest was the choice of band method used with students. A replication of this study with students who use a more traditional band method book may also produce different results. Therefore, further self-evaluation methodology may be a practice strategy toward improving students' self-rating accuracy and independent musicianship.

REFERENCES

- Abeles, H. F. (1973, Fall). Development and validation of a clarinet performance adjudication scale. *Journal of Research in Music Education*, 21(3), 246-255.
- Aitchison, R. A. (1995). The effects of self-evaluation techniques on the musical performance, self-evaluation accuracy, motivation, and self-esteem of middle school instrumental music students. (Doctoral dissertation, University of Iowa). *Dissertation Abstracts International*, 56-10A, 3875.
- Bergee, M. J. (1987). Ringing in the changes: General John Eaton and the 1886 public school music survey. *Journal of Research in Music Education*, 35(2), 103-116.
- Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education*, 41, 19-27.
- Bergee, M. J. & Roberts, L.C. (2002). Effects of small-group peer interaction on self-evaluation music performance. *Journal of Research in Music Education*, 50(3), 256-268.
- Bundy, O. R. (1987). Instrumentalists' perception of their performance as measured by detection of pitch and rhythm errors under live and recorded conditions. (Doctoral dissertation, Pennsylvania State University). *Dissertation Abstracts International*, 48-10A, 2567.
- Brown, K. J. (1999). What kind of text – for whom and when? Textual scaffolding for beginning readers. *The Reading Teacher*, 4, 292-307.
- Byo, J. L. (1990). Recognition of intensity contrasts in gestures of beginning conductors. *Journal of Research in Music Education*, 38, 157-163.

- Byo, J. L., & Brooks, R. (1994). A comparison of junior high musicians' and music educators' performance evaluations of instrumental music. *Contributions to Music Education, 21*, 26-38.
- Colwell, C. M. (1995). Effects of teaching setting and self-evaluation on teacher intensity behaviors. *Journal of Research in Music Education, 43*(1), 6-21.
- Cooksey, J. M. (1977). A facet-factorial approach to rating high school choral music performance. *Journal of Research in Music Education, 25*(2), 100-114.
- Covington, M. V. & Roberts, B. (1994). Self worth and college achievement: Motivational and personality correlates. In P. R. Pintrich, D. R. Brown, & C. E. Weinstein (Eds.), *Student motivation, cognition, and learning: Essays in honor of Wilbert J. McKeachie*, 157-187. Hillsdale, NJ: Erlbaum.
- Davidson, L., & Scripp, L. (1992). Surveying the coordinates of cognitive skills in music. In R. Colwell (Ed.), *Handbook of research on music teaching and learning* (pp. 392-413). New York: Schirmer.
- Davis, L. P. (1981). The effects of structured singing activities and evaluation practice on elementary band students' instrumental achievement. (Doctoral dissertation, Ohio State University). *Dissertation Abstracts International, 42-07A*, 3051.
- Elliott, D. J. (1995). *Music matters: A new philosophy of music education*. New York: Oxford University Press.
- Ericsson, K. A. (1997). Deliberate practice and the acquisition of expert performance: An overview. In H. Jorgensen & A. C. Lehman (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 9-52). Oslo, Norway: Norges musikkholshole.

- Eshel, Y., & Klein, Z. (1981). Development of academic self-concept of lower class and middle class primary school children. *Journal of Educational Psychology, 73*, 287-293.
- Fiske, H. E. (1975, Fall). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education, 23*(3), 186-196.
- Fiske, H. E. (1978). The effect of training procedure in musical performance evaluation on judge reliability. *Ontario Education Research Council Report*.
- Gordon, E. (2002). *Rating scales and their uses for measuring and evaluating achievement in music performance*. Chicago: GIA.
- Gutsch, K. U. (1964). One approach toward the development of an individual test for assessing one aspect of instrumental music achievement. *Bulletin of the Council for Research in Music Education, 2*(2), 1-5.
- Hewitt, M. P. (2001). The effects of self-evaluation, self-listening, and modeling on junior high instrumentalists' music performance and practice attitude. *Journal of Research in Music Education, 49*(4), 307-322.
- Hewitt, M. P. (2002). Self-evaluation tendencies of junior high instrumentalists. *Journal of Research in Music Education, 50*(3), 215-226.
- Jorgensen, H. (1995). Teaching and learning strategies in instrumental practice: A report on research in progress. In J. A. Taylor (Ed.), *Transatlantic roads of music education: World Views*, 47-51. Tallahassee, FL: Center for Music Research.

- Kepner, C. B. (1986). The effects of performance familiarity, listening condition, and type of performance error on correctness of performance error detection by 50 high school instrumentalists as explained through a sensory blocking theory. (Doctoral dissertation, Kent State University, Kent, OH). *Dissertation Abstracts International, 47-05A*, 1643.
- Kidd, R. L. (1975). *The construction and validation of a scale of trombone performance skills*. Unpublished doctoral dissertation, University of Illinois at Urbana Champagne, 1975.
- Linn, R., & Miller, M. D. (2005) *Measurement and assessment in teaching (9th Edition)*. Upper Saddle River, NJ: Prentice-Hall.
- McPherson, G. E., & Zimmerman, B. J. (2002). Self-regulation of musical learning: A social cognitive perspective. In R. Colwell & C. Richardson (Eds.), *The new handbook of research on music teaching and learning*, 327-347. New York: Oxford University Press.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice and performance. *Psychological Review, 91*, 328-346.
- Nicholls, J. G., & Miller, A. T. (1983). Reasoning about the ability of self and others: A developmental study. *Child Development, 55*, 1990-1999.
- Rosenthal, R. K. (1985). Improving teacher effectiveness through self-assessment: A case study. *Update: The Applications of Research in Music Education, 3(2)*, 17-21.
- Saunders, T. C., & Holohan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *Journal of Research in Music Education, 45*, 259-272.

- Slavin, R. E. (1991). *Educational psychology* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Sparks, G. E. (1990). The effect of self-evaluation on musical achievement, attentiveness, and attitudes of elementary school instrumental students. (Doctoral dissertation, Louisiana State University). *Dissertation Abstracts International*, 51-09A, 3009.
- Stivers, J. D. (1972). A reliability and validity study of the Watkins-Farnum performance scale (Unpublished doctoral dissertation, University of Illinois at Urbana Champagne, 1972). *Dissertation Abstracts International*, 34-O2A, 815-816A.
- Watkins, J. G., & Farnum, S. E. (1954). *The Watkins-Farnum Performance Scale*. Milwaukee, WI: Hal Leonard Publishing.
- Yarbrough, C., Wapnick, J., & Kelly, R. (1979). Effect of videotape feedback techniques on performance, verbalization, and attitude of beginning conductors. *Journal of Research in Music Education*, 27(2), 103-112.
- Zdzinski, S. F. (1991, Summer). Measurement of solo instrumental music performance: A review of literature. *Bulletin of the Council for Research in Music Education*, (109), 47-58.
- Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*, 13-39. San Diego, CA: Academic Press.