Texas Music Education Research 2008

Reports of Research in Music Education
Presented at the Annual Meeting of the Texas Music Educators Association
San Antonio, Texas, February, 2008

Robert A. Duke, Chair TMEA Research Committee
School of Music, The University of Texas at Austin

Edited by: Edited by Mary Ellen Cavitt,
Texas State University, San Marcos

Published by the Texas Music Educators Association, Austin, Texas

CONTENTS

Edited by
Mary Ellen Cavitt, Texas State University—San Marcos

The Effect of Self-Directed Peer Teaching on Undergraduate Acquisition of Specified Music Teaching Skills

Janice N. Killian, Keith G. Dye, and Jeremy J. Buckner
Texas Tech University

Peer teaching is a time-honored method of practicing teaching skills prior to entering a classroom. Traditionally, a pre-service educator teaches an assigned lesson and the instructor gives feedback regarding lesson content and structure (Fuller & Brown, 1975). Drawbacks of peer teaching include the fact that peers often can perform the content of the lesson; thus the pre-service teacher does not get accurate feedback nor does he/she practice error-detection skills because the "students" do not make errors on familiar material. Additionally, when the instructor gives feedback, the pre-service teacher may enter the mode of "What did I get?" rather than pursuing self-evaluatory paths that frequently will be the only source of feedback when he/she actually teaches in the field.

Will pre-service teachers improve their teaching skills and their ability to accurately evaluate themselves when given the opportunity to teach their peers unfamiliar material and to rely on their own powers of self-evaluation rather than those of the instructor? Following the teaching models suggested by Paul (1998), Paul, Teachout, Sullivan, Kelly, Bauer and Raiber (2001) and Raiber (2005, 2006), this study was designed to address the question by examining the effects of self directed evaluation of peer teaching episodes on the acquisition of observable teaching skills among undergraduate pre-service music educators. A particular focus of this preliminary analysis involved participant reflections and attitudes about their perceived growth as young educators.

Method

Participants consisted of senior music education majors (N=45) enrolled at a large southwestern school of music. Each had completed two semesters both of conducting and elementary music methods and were currently enrolled in either band (n=24), choir (n=12) or orchestra (n=9) methods classes as well as the single pre-student teaching music education course that was the focus of this investigation.

Students were assigned four peer teaching episodes as part of their course. In order to insure that they were teaching material unfamiliar to their peers, but highly familiar to themselves, students composed new compositions for each teaching episode (TE). Each teaching episode

was designed to be progressively more complex and to require increased levels of musical skill. Thus, instructions for the teaching episodes included the following:

TE 1: Compose and teach a 16-measure, 4-part rhythmic composition
TE 2: Re-teach TE 1 composition or a new 16-measure, 4-part rhythmic composition
TE 3: Compose and teach a 32-measure vocal melody with text
TE 4: Arrange and teach an assigned Bach chorale for wind instruments represented in the class (band majors), or for SATB choir with orchestra (choral and orchestral majors).

The class met as a whole for traditional class content. For the teaching episodes, however, students were divided into three smaller groups (n=15) allowing time for informal peer and instructor verbal feedback following each teach. Students were randomly assigned to three different peer-teaching sites under the guidance of three instructors for each teaching episode. Thus an individual student taught each of the four episodes to a different configuration of his/her peers. Teaching episodes ranged in length from 8-12 minutes and were recorded for subsequent self-analysis and reflection using Sony 105 Mini-DVD camcorders.

Students submitted a structured teaching plan electronically prior to each teaching episode (TE). Anyone failing to submit a plan was not permitted to teach that episode. Students received instructor feedback on their plan and completed a guided reflection after viewing their teaching DVD. Planning and reflection forms (see Figures 1 and 2) were adapted from Raiber (2005, 2006).

MUED 3311
Planning Framework Name: Teaching Episode Date:

Type your plan in the form below. Boxes will expand as you type in them.

Teaching Episode: 1 2 3 4

Rehearsal Objective: (What will your students be able to do as a result of this rehearsal **with you**?)

Assessment of the Rehearsal: (How will **you** know if your **students can do** what they need to?)

Rehearsal Plan: (A step by step plan for the rehearsal - with timing - that leads to your objective.)

Time	Activity	Description	Purpose of Activity	Assessment

Figure 1. Lesson Plan Form Adapted from Raiber (2005, 2006).

Teaching Episode 1_____ 2_____ 3_____ 4_____ Name:

Based upon what you see in your video…

What is your overall impression of this rehearsal? What leads you to this conclusion?

How well did your planning work to help you achieve your goal(s) for this rehearsal? Is there anything that you would do differently had you the opportunity to plan this lesson again?

How effective were your delivery skills? Is there anything that you need to improve upon to be more effective? How do you plan to work on this?

How was the pacing in your rehearsal? Were all your teaching cycles complete? Did your feedback always match your set? What might you do in future teaching to make this more effective?

What rehearsal technique(s) was/were most effective in this rehearsal? Why do you think so? How might you be able to apply this to other rehearsals?

What rehearsal technique(s) was/were least effective in this rehearsal? Why do you think so? How might you be able to fix this in other rehearsals?

What conducting technique(s) was/were most effective in this rehearsal? Why do you think so? How might you be able to apply this to other rehearsals?

What conducting technique(s) was/were least effective in this rehearsal? Why do you think so? How might you be able to fix this in other rehearsals?

Based upon what you hear in your rehearsal recording…

What is your general impression of the core sound in the ensemble? Why? What needs to be done to either maintain this sound or improve upon it?

What area needs the most attention in your next rehearsal? Why do you think so? What can you do to fix this?

Figure 2. Rehearsal Analysis/Reflection Form Adapted from Raiber (2005, 2006).

Instructors (who were also the researchers for this study) monitored submission of plans and reflections, and guided class discussion following each teaching episode; however, the quality of the teaching was not graded. Thus, students self-evaluated the quality of their own teaching, and observed their peers' teaching techniques while serving as participants in each teaching episode. Following the final plan/teach/reflect rotation, students completed a survey regarding their perceived success as a teacher, their opinion about personal learning and growth regarding teaching and self-evaluation, and their preference for this type of class procedure.

Results

Data consisted of survey responses (both Likert scale and free response answers) from the 43 students who completed the survey. Additionally, for the purposes of this study, the videos of eight students were selected for more in depth analysis. This non-random select sample of eight students was chosen by the three course instructors as representative of a variety of teaching skill levels, allowing us to evaluate how the plan/teach/evaluate model worked with differing students. The eight students included band (*n*= 4), choir (*n*=2) and orchestra (*n*=2), representing approximately the percentage of each specialization within the larger group.

Results

Results of the survey (Tables 1 and 2) were tabulated and displayed in terms of mean responses.

Table 1
Peer Teaching Survey Responses (n=43)

• Time spent in preparation of:

	0-30 min	31-60 min	61-90 min	91+min	
Materials	*5*	*3*	*5*	*30*	*M = 3.40 / 4.0*
Plan form	*15*	*16*	*2*	*10*	*M = 2.16 / 4.0*
Reflection form	*20*	*21*	*1*	*1*	*M = 1.60 / 4.0*
Individual Prep.	*15*	*13*	*3*	*10*	*M = 2.20 / 4.0*
				*(*2 no response)*	

• Number of times recording was viewed prior to completing reflection:

Never	1	2	3	3+	
3	*17*	*20*	*2*	*1*	*M = 2.56 / 5.0*

• To what degree was planning helpful?

Not	Somewhat	Helpful	Extremely	
2	*6*	*19*	*16*	*M = 3.14 / 4.0*

• To what degree did you stick to your plan?

Not	Less than ½ time	More than ½ time	Almost all time	
0	*11*	*21*	*11*	*M = 3.00 / 4.0*

• Not grading Teaching Episodes helpful?

Yes	No	No Opinion
29	*3*	*11*

• More or less instructor feedback wanted?

More	Less	Same
19	*0*	*24*

• Teaching enhanced by using original materials?

Yes	No	No Opinion	
25	*6*	*11*	*(*1 no response)*

• Would you have preferred class to be traditional lecture/notes/test?

Yes	No	No Opinion
0	*42*	*1*

• Do you feel that overall your teaching has improved this semester?

Yes	No	No Opinion
43	*0*	*0*

Table 2
Peer Teaching Survey Responses (n=43)

• How much did your teaching improve in the following categories?

	1 (no)	2 (little)	3 (no need)	4 (some)	5 (great)	
Confidence	*0*	*1*	*1*	*10*	*31*	*M = 4.65 / 5.0*
Delivery	*0*	*0*	*0*	*23*	*20*	*M = 4.47*
Leadership	*0*	*0*	*3*	*22*	*18*	*M = 4.35*
Planning	*0*	*3*	*4*	*18*	*18*	*M = 4.19*
Clarity of directions	*0*	*3*	*2*	*23*	*15*	*M = 4.16*
Use of time	*0*	*2*	*6*	*19*	*16*	*M = 4.40*
Eye contact	*0*	*2*	*6*	*25*	*10*	*M = 4.00*
Proximity	*2*	*7*	*5*	*21*	*8*	*M = 3.60*
Conducting	*2*	*13*	*9*	*13*	*6*	*M = 3.19*
Other	*0*	*0*	*0*	*0*	*5**	*M = 5.00*

*E.g., Teaching musical elements, questioning, not feeling like an idiot, relating to students, relationships with peers

• How valuable have the following Teaching Episode (TE) activities been?

	1 (none)	2 (little)	3 (some)	4 (high)	
Teaching	*0*	*0*	*2*	*41*	*M = 3.95 / 4.0*
Planning	*0*	*4*	*13*	*26*	*M = 3.51*
Participation & Observation	*0*	*3*	*16*	*24*	*M = 3.49*
Reflection	*3*	*5*	*16*	*19*	*M = 3.19*
Composition	*2*	*6*	*24*	*11*	*M = 3.02*

In addition, several survey questions requested free response reactions. Written responses were read and categorized for consistency of comments. Free response questions included: Would you prefer that the class be in a traditional lecture/notes/test format? All but one respondent preferred the peer teaching. The following categories consistently appeared in participants' responses.

Learning by doing: "I do not think we would have learned as much. It is fine to talk about what to do, but there is no substitute for actually doing it." "You can't learn to teach without actually teaching."

Personal Learning Style: "I do not learn as well by taking notes and hearing lecture." "I learn a lot better by actually doing things. " "These teaching projects were especially helpful in identifying my strengths and weaknesses. I would not have gotten that from a lecture class."

Gained confidence: "I feel more confident in teaching music thanks to my experiences in this class." "The TEs helped me and others build confidence & actually gave us a chance to teach."

Do you feel that overall your teaching has improved this semester? Respondents unanimously agreed their teaching had improved. Consistent response categories included:

Confidence: Comfort and confidence were mentioned by 24 of the 43 students. "I don't get nervous anymore and feel more confident in front of the group." "I have a better idea about what works and what doesn't as well as more confidence in front of a group."
Planning: "I have been able to gain strategies for teaching within a music classroom. This idea, especially lesson planning for an arts classroom, was out of my realm of experience before this semester." "I have learned a lot by planning lessons and actually implementing those plans."
Personal Development: "Starting to find who I am as a teacher vs. student vs. person." "Because teaching is all about taking personal responsibility for your mistakes and learning with the students, for the students."

How valuable have the following TE activities been for you?

Specific comments mirrored the statistical means reported earlier (see Table 2). Consistent categories of responses included the following:

Teaching

The vast majority (n = 41) felt that teaching was a "high value" learning experience while 2 rated teaching as "some value" (M = 3.95 out of 4.0). Planning (M = 3.51) was rated as "high value" by 26, "some value" by 13 and "less value" by 4. Representative comments included: "Each of the activities sharpened my skills as a teacher of music, specifically the planning and reflection portions." "Honestly, I didn't like doing plannings for each teaching. I didn't follow them because there is no way of knowing how the class will go."

Class Participation and Observation of Peers

Most (M = 3.49) "highly valued" the experience. Representative comments included: "I believe class participation and peer observation made me think outside the box in terms of how I run a rehearsal." "I learned more from teaching myself than watching others."

Reflection

Responses on reflection were more mixed (M = 3.19) but generally positive. Representative comments included: "It really helped to see myself on video and analyze myself." "I enjoyed watching my videos for feedback, but articulating my feelings was difficult" "Reflection was hard because it is hard to watch yourself analytically and not critically." "I knew how I did – reflection seemed redundant." "Reflecting caused more of a sense of self-doubt."

Composition

The lowest mean score (3.02 out of 4 points) was chosen; however, the reactions were still generally positive. Free responses on composition indicted differences in understanding why original compositions were assigned. "Composition isn't an area that I need to work on." "Being able to compose and teach my own work made me more apt to do it--I wanted it to be good." "I

had never really composed before and I liked it." "Serious composition work I need improvement but not really a rhythmic motive."

The survey also asked individuals to rate how much they believed their teaching improved on specified categories (see Table 1). Note that these categories were not assigned as a focus of the class, but comments about these categories did arise during informal peer and instructor feedback following each teaching episode. Greatest personal improvement was noted in the areas of "Confidence," "Delivery," "Leadership," and "Planning." Least improvement appeared in "Conducting."

Analysis of Video Recordings of Teaching Episodes (TEs)

This stage of our continuing research was designed to identify teaching behaviors that might have improved based on self-evaluation in the absence of instruction. Preliminary comparisons of TE 1 (rhythmic composition) and TE 4 (Bach chorale) indicated striking differences in teacher talk time and musical performance time between the two teaching episodes which we speculated were due to the differences in the material being taught. Thus, analysis for this preliminary stage of our research was limited to comparisons of Teaching Episode 1 and 2 because both involved the same task and the same material (16 measure 4-voice rhythmic composition), and differences in musical factors between teaching episodes could be lessened. Additionally, a 2-minute excerpt from each teaching episode (approximately 1 minute into a 8-12 minute teach) was isolated to allow a snapshot of the teacher after initial instructions had been given. Thus 2 minute excerpts (ranging from 1 minute 53 seconds to 2 minutes 3 seconds in length) for 8 students across 2 teaching episodes were analyzed yielding 16 two-minute excerpts in all.

Selected videos were analyzed using SCRIBE (Simple Computer Recording Interface for Behavioral Evaluation) 3.0a4 software (Duke & Farra, 2001), which allows researchers to select what categories they wish to observe and then, using multiple passes over the same video, measure how often or for how long the chosen categories occurred. The categories chosen (Siebenaler, 1997) included timed events (teacher talk, students perform, students talk, teacher performs) as well as frequency events (teacher talks while students perform, teacher instruction, teacher questions students, students respond, teacher models, general verbal feedback, specific verbal feedback, teacher corrects or extends, and off task teacher comment). Figures 3 and 4 display samples of the SCRIBE template and the resulting timeline.

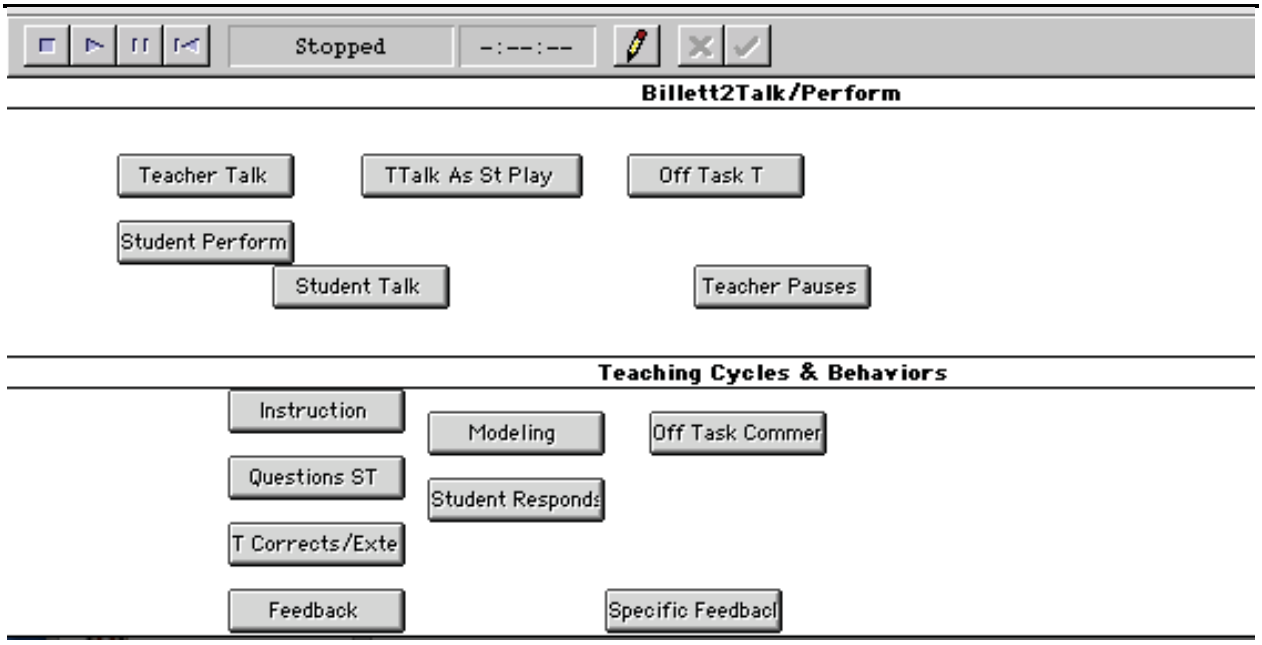

Figure 3. Template of SCRIBE data collection.

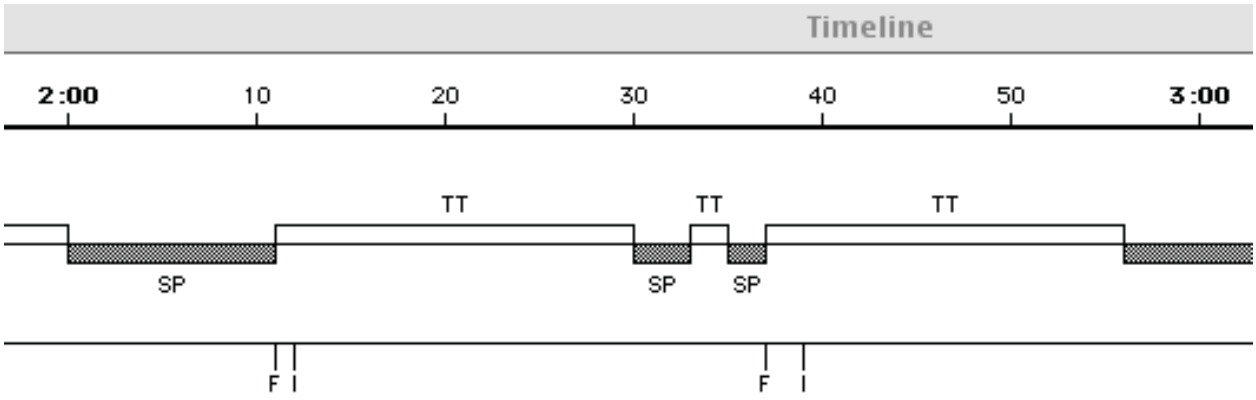

Figure 4. *Sample of SCRIBE timeline for minute 2 through minute 3 (TT = Teacher Talk, SP = Student Perform, F= General Feedback, I = Instruction)*

Of particular interest was the amount of overall teacher interaction with students (Madsen, 2003; Madsen, Standley, Byo & Cassidy, 1992). Teacher interaction was operationally defined as the total number of categories present in the 2 minutes, a figure that is compiled by the SCRIBE software. Thus, a teacher who either gave instructions continuously or asked the students to perform continuously would have relatively few interactions. A teacher who gave short instructions and asked students to respond quickly (e.g., echo clapping) would have many more interactions (see Figure 4). It was predicted that the better teachers would exhibit a higher percentage of overall interactions (Duke, Prickett & Jellison, 1998), and that the second teaching episode would indicate an increase in overall interactions. Overall, during the 2-minute excerpts teachers interacted with their students between 9 and 54 times. Tables 3 and 4 shows the results of the teacher characteristics gathered for this study and indicates that 5 of the 8 teachers increased interactions between TE 1 and TE 2.

Table 3
Interactions, Teacher Talk, and Student Perform on Teaching Episodes (TEs) 1 and 2.

Teacher (beginning interaction-ending interaction)	TE #	Total Time	Inter-actions/ Seconds	Inter-actions %	Teacher Talk % of Time	Student Perform % of Time
1. Band						
(24-59)	1	2:02	36/122	29.5%	85.2	16.4
(17-42)	2	1:53	26/113	23.0%	83.2	23.9
2. Band						
(6-59)	1	2:00	54/120	45.1%	77.5	0.0
(10-24)	2	1:55	15/115	13.0%	32.2	16.5
3. Choir						
(11-26)	1	1:53	16/113	14.2%	24.8	74.3
(16-35)	2	1:55	20/115	17.4%	59.1	52.2
4. Orchestra						
(7-27)	1	2:01	21/121	17.4%	40.5	57.9
(3-40)	2	1:57	34/117	29.1%	65.0	29.9
5. Orchestra						
(17-68)	1	2:03	52/123	42.3%	61.8	19.5
(12-58)	2	1:59	47/119	39.5%	77.3	15.1
6. Band						
(8-17)	1	1:54	9/114	07.9%	33.3	56.1
(3-26)	2	2:04	24/124	19.4%	55.6	43.5
7. Band						
(23-41)	1	1:59	19/119	16.0%	70.6	29.4
(24-54)	2	2:13	31/133	23.3%	54.1	21.1
8. Choir						
(3-23)	1	2:00	21/120	17.5%	57.5	31.7
(3-29)	2	1:55	27/115	23.5%	64.3	31.3

Table 4

Frequency of Behaviors Observed During TE 1 and 2

Teacher (beginning interaction-ending interaction)	TE #	Teacher Verbal Feed back	Teacher Questions Students	Teacher Talks as Students Perform	Student Talk	Teacher Instructs	Teacher Models	Teacher Corrects/ Extends
1. Band								
(24-59)	1	6	0	0	1	8	0	3
(17-42)	2	4	1	0	2	6	0	0
2. Band								
(6-59)	1	3	7	0	10	8	3	1
(10-24)	2	3	1	2	0	3	0	0
3. Choir								
(11-26)	1	1	0	0	0	4	1	2
(16-35)	2	4	0	0	0	5	1	1
4. Orchestra								
(7-27)	1	2	4	0	2	4	0	0
(3-40)	2	9	1	0	2	9	0	1
5. Orchestra								
(17-68)	1	0	0	0	0	10	11	0
(12-58)	2	8	1	2	1	11	7	0
6. Band								
(8-17)	1	1	0	0	0	4	0	2
(3-26)	2	2	0	0	1	4	0	3
7. Band								
(23-41)	1	3	0	0	0	6	2	1
(24-54)	2	2	0	0	1	7	5	2
8. Choir								
(3-23)	1	4	0	0	0	5	3	2
(3-29)	2	3	1	0	2	7	1	3

Similarly, since data indicates that experienced teachers talk less than inexperienced ones (Goolsby, 1996; Madsen, 2003), we expected teachers to decrease the amount of teacher talk by asking their students to become more involved in the lesson between TE 1 and TE 2. Table 2 indicates that a decrease in teacher talk was in evidence for three teachers. Five of the teachers showed a marked increase in the amount of verbal feedback offered to students. No pattern appeared to develop in the frequency with which teachers questioned their students, at least not during the two-minute excerpts we examined.

Discussion

Results of this study indicated that:

1) Pre-service teachers unanimously preferred the use of peer teaching episodes to lecture classes and believed they learned much about being a teacher using these procedures.
2) Pre-service teachers indicated that they improved the most on the following skills or attributes: confidence, delivery, leadership, planning, clarity of directions, use of time, eye contact, use of proximity, and conducting.
3) Pre-service teachers particularly valued the opportunity to teach, plan, and participate/observe their peers. They provided less positive ratings to the value of composing their own music to teach their peers and to the process of written reflections, although their rankings were still generally positive.
4) Detailed analyses of the video teaching samples revealed that although individual teachers exhibited definite improvement on the categories observed (particularly in the areas of overall interaction and less teacher talk), much further research is necessary before observable and generalizable statements about factors affecting teaching quality are possible.

Many of the findings of this study were consistent with previous research in both music and general education. It is interesting to note that less emphasis was on musical skills (e.g., conducting) than on social skills (e.g., confidence), which is consistent with the opinion of Lee Shulman, President of the Carnegie Foundation for the Advancement of Teaching (Tell, 2001). Similar to participants in both music (Bowers, 2001; Colwell, 1995) and non-music settings (Anderson & Graebell, 1990), our participants valued experiential learning and felt they learned more under peer teaching conditions. Also consistent with previous research among music educators in which "confidence" was listed as the most important of nine items by both experienced and inexperienced teachers (Teachout, 1997), our pre-service teachers mentioned gains in confidence as the most valued part of the experience. Perhaps this improvement in confidence and respondents' beliefs in the overall value of experiential peer teaching are the most important factors we noticed upon preliminary analysis.

We plan to continue examining these procedures while addressing an important question: Are students able to transfer gains in confidence, delivery, planning and leadership into structured field based experiences in which they teach rather than merely observe? The second semester of this course will allow us to examine this question for these same students.

Self-perceived student gains were primarily in the areas of confidence, planning, delivery and leadership, but we were unable to identify consistent improvement in specified music teaching skills. Perhaps we were not looking at the right skills. Will both musical teaching skills as well as social skills increase if students are given focused instruction? The fact that 19 students would have preferred more instructor feedback while none requested less (see Table 1) would seem to indicate that they would value more instruction. In this preliminary study, students viewed videos of their teaching and reflected on those videos; they were not shown the SCRIBE data that we collected regarding their videos. It might be speculated that more focused viewing would yield benefits. For example, will specific skills increase if students tabulate or at least view their own SCRIBE data? Which teaching skills might be identified as most important? Is there an optimal sequence in which to present those skills to novice music educators?

Our students unanimously preferred experiential learning to traditional classes, which raises questions regarding the balance between undergraduate self-reflection and instructor input. We plan to continue this process in a databased manner in an ongoing effort to better prepare undergraduate music educators.

References

Anderson, N. & Graebell, L (1990). Usefulness of an early field experience. *The Teacher Educator, 26,* 13-20.

Bowers, J. (2001). A field experience partnership for teacher education with university, public school and community participants. *Bulletin of the Council for Research in Music Education, 148,* 3-11.

Colwell, C. (1995). Effect of teaching setting and self-evaluation on teacher intensity behaviors. *Journal of Research in Music Education, 43,* 6-21.

Duke, R., Prickett, C. & Jellison, J. (1998). Empirical description of the pace of music instruction. *Journal of Research in Music Education, 46,* 265-280.

Duke, R., & Farra, Y. (2001). *SCRIBE: Simple Computer Recording Interface for Behavioral Evaluation.* Version 3.0a4. Center for Music Learning. The University of Texas at Austin.

Fuller, F., & Brown, O. (1975). Becoming a teacher. In K. Ryan (Ed.). *Teacher education, Pt. II: 74th yearbook of the national society for the study of education.* University of Chicago.

Goolsby, T. (1996). Time use in instrumental rehearsals: A comparison of experienced, novice, and student teachers. *Journal of Research in Music Education, 44,* 286-303.

Madsen, C., Standley, J., Byo, J. & Cassidy, J. (1992). Assessment of effective teaching by instrumental music student teachers and experts. *UPDATE: Applications of Resarch in Music Education, 10* 20-24.

Madsen, C. (2003). The effect of accuracy of instruction, teacher delivery, and student attentiveness on musicians' evaluation of teacher effectiveness. *Journal of Research in Music Education, 51,* 38-50.

Paul, S. J. (1998). The effects of peer teaching experiences on the professional teacher role development of undergraduate instrumental music education majors. *Bulletin of the Council for Research in Music Education, 137,* 73-92.

Paul, S. J., Teachout, D. J., Sullivan, J. M., Kelly, S. N., Bauer W. I., and Raiber, M. A. (2001). Authentic context learning activities in instrumental music. *Journal of Research in Music Education, 49,* 136-145.

Raiber, M. (2005). *The emerging instrumental music educator: Developing the professional and occupational roles*. Best Practice Paper. Society for Music Teacher Education Symposium. University of North Carolina at Greensboro.

Raiber, M. (2006). Invited presentation at Texas Tech University. Lubbock, Texas.

Siebenaler, D. (1997). Analysis of teacher-student interactions in the piano lessons of adults and children. *Journal of Research in Music Education, 45,* 6-20.

Teachout, D. (1997). Preservice & experienced teacher opinions of skills & behaviors important to successful music teaching. *Journal of Research in Music Education, 45*, 41-50.

Tell, C. (2001). Appreciating good teaching: A conversation with Lee Shulman. *Educational Leadership, 58*, 6-11.

Edited by
Mary Ellen Cavitt, Texas State University—San Marcos

The Effects of Computer-Augmented Feedback on Self-Evaluation Skills of Student Musicians

Melanie J. McKown
Boston University
Mary Ellen Cavitt
Texas State University—San Marcos

The acquisition of any new skill requires practice. Whether the skill is a foreign language, practical application of a math concept, or the ability to play an instrument, most students require repetition of the new skill accompanied by some form of feedback, whether intrinsic or external, until mastery is achieved.

As instructional methodologies and goals in the arts are developed and refined, it is only natural that our assessment methods also be examined. Byo (1998) suggested adapting an assessment philosophy that not only provides evidence of progress but opportunities to teach. Assessments should not just be information-gathering sessions, but allow for growth of the individual. Designing assessments that are able to deliver useful feedback within an acceptable time frame should be a high priority for the educator.

Colwell (2002) suggested that performing musicians should constantly engage in self-evaluation in order to hone their technique. Music teachers do not normally require self-evaluation techniques of their students, which often creates students that are unable to guide themselves through successful study of an independent work and who are totally reliant on their teachers for assessment of their performance. Time restrictions in the classroom can leave the students without any real concept of how they are performing or what needs improvement.

An important element of self-evaluation is comparison of the performance to a standard, model or goal, and musicians need a clear understanding of the aural concept they are trying to emulate before they can accurately make an evaluation (Hewitt, 2001). When combined with self-evaluation, listening to aural models seems to be more effective than not using a model, especially in the areas of tone, rhythm, interpretation, and overall performance (Hewitt). Whether listening to themselves or others, students should strive to obtain the knowledge necessary to make judgments based on appropriate factors, not egocentric or subjective motives.

Colwell also suggested that students should judge themselves and each other as an effective means of evaluation (Colwell & Goolsby, 1992). However, it should be remembered that this technique is only effective to the extent that students are able to accurately judge one's own performance. Peer assessments are generally higher than teacher assessments, and small-group interaction does not have a lasting effect on self-evaluation skills (Bergee & Cecconi-Roberts,

2002). Conversely, motor skills studies have shown that having students evaluate others' performances during the interval immediately following their performance and before receiving feedback (KR-Delay Interval) may hinder their own subsequent learning (Swinnen, 1990). The music teacher, therefore, should give careful consideration to the use of self-evaluation and peer-evaluation during practice in a group setting. Timing of delivery of feedback for each student's performance should be adjusted so that engaging in peer-evaluation does not have a deleterious effect on learning. Determining whether students are truly able to properly evaluate their own performance must be considered before relying heavily on self-evaluation without some type of guided feedback.

Hewitt (2002) investigated the ability of junior high students to self-evaluate their performances both with and without the use of a model. Numerous studies have previously indicated that use of a model improved musical performance (Dickey, 1991, 1992; Henley, 2001; Hewitt, 2001; Linklater, 1997; Puopolo, 1971; Rosenthal, 1984; Rosenthal, Wilson, Evans, & Greenwalt, 1988; Zurcher, 1975), although a study by Anderson (1981) determined that tape-recorded aural models did not have an observable effect. Hewitt's results showed that the use of a model as a comparison had no effect on junior high students' ability to self-evaluate effectively, and concluded that they may not be able to self-evaluate. Most studies regarding this topic indicate that students are inconsistent in their evaluations of their own performances (Atchison, 1995; Byo & Brooks, 1994; Hewitt, 2002, 2005) although Hewitt (2005) did find that high school students may be more accurate than middle school students in areas other than melody and rhythm, which were found to be similar.

Audio and video recorders, when used in the classroom setting, can be an excellent tool for both the teacher and the student to self-evaluate and alter actions and habits accordingly. Kohut (1985) reminded us of the effectiveness of recording as an evaluation after performance and suggested incorporating recordings into the rehearsal as instant aural feedback. While these tools may be effective, the time used to record and review the tapes during class is often limited. Frequently, music educators are rushed during class time to evaluate large numbers of students and are unable to devote adequate time to listening to individuals perform. A common assessment practice, therefore, makes use of recording or video equipment, having students tape the performance for later evaluation (Byo, 2001; Garafalo, 1995; Gonzo & Forsythe, 1976; Young & Regenberg, 2001). While this eases the burden created by lack of contact time to evaluate, it precludes the students from receiving meaningful augmented feedback immediately after their performance. With this system, students are often only ranked after an evaluation, with no indication of what was wrong or right in their performance. The student sees only the final result in the assessment process, the assignment of a grade.

Young and Regenberg (2001) suggested that by reviewing recordings of their own performances, students were able to objectively evaluate their abilities and more closely align their perception of their performance and the reality of the performance. Colwell suggested that computers be used to judge pitch and rhythm accurately and with precision, providing a safe environment in which the student can perform and receive feedback without risk (Colwell & Goolsby, 1992).

In *Motor Learning: Concepts and Applications,* Magill (2004) described two types of feedback: 1) *Task-intrinsic feedback* comes from within the performer, and is the sensory-perceptual information that is part of the performance; and 2) *Augmented feedback* is generally information that comes from outside the performer, that which may be used to supplement the internal, performance-related information. Any feedback that is not internal to the performer is considered augmented feedback, regardless of its content or depth. This feedback comes in

various forms, such as teacher corrections, videotape, or computer generated graphics. Augmented feedback may be further divided into two groups: (1) *Knowledge of Results (KR),* which is external information about the outcome of the performance; and (2) *Knowledge of Performance (KP),* external information about the movement characteristics that led to the performance outcome. The time between performance attempts is known as the *inter-trial delay,* and may be further divided as (a) *KR-delay Interval,* the time between a performance attempt and receiving feedback, and (b) *Post-KR Interval,* the time after receiving KR and before attempting the performance again.

Colwell (1970) asserted that a major value of testing and measuring is in the feedback, but that too often, grades are received at the end of the term. These "postmortem acts" often provide no feedback to the student other than the assignment of a grade, which is often not at all useful to the student. Colwell suggested consideration of computer-assisted instruction as a means to minimize grades while preserving evaluation, which is critical to learning (Colwell, 2002).

Computer-assisted instruction (CAI) has been a key use of technology for many years and software developers continue to look for areas that easily adapt to its use. Many popular instrumental methods now include accompaniment software that allows the student to manipulate tempo, play the melody as a model and record their performances. As technology has evolved, publishers have moved away from the traditionally included fixed tempo audio recordings and moved to computer programs that allow for greater control by the performer. Many also include practice aids such as built-in fingering charts for the different instruments and can transpose exercises into appropriate keys.

As we incorporate technology and more creative approaches to our curriculum, it is only natural that we also move away from standard paper-and-pencil tests and look for different ways to assess our students (Hickey & Webster, 2001). Sebald (2000) stated that automated assessment and reporting has freed the educator and made frequent feedback to all students a possibility. This high level of feedback frequency both from the teacher and from the test results was previously unattainable, but now it enables more effective progress monitoring (Sebald, 2000). The availability of immediate aural feedback in conjunction with guided learning can be beneficial to student acquisition of skills (Hoffmann, 1991). While educators embrace the use of technology in their classrooms as supplements to the curriculum, little research has been done on the development and implementation of assessment modules for use in music performance training. Most studies have focused on the use of computers as an assessment tool in the gauging of components of music aptitude: (1) pitch awareness (Brick, 1984; Peters, 1993); (2) ability to detect changes in melody, rhythm, texture and tonality (Venn, 1990); (3) rhythm performance assessment (Meeuwsen, Flohr, & Fink, 1998); and (4) evaluation of music ability (Hickey & Webster, 1999).

Because of the inclusion of assessment modules in many method books at no additional cost, access to some form of computer-assisted instruction in music is now available to virtually all students. Programs also available for free download may be combined with recent commercial notation program features to create an authoring/assessment combination which allows educators the ability to assess rhythm and pitch in any musical passage they choose, and to offer yet another approach to delivering feedback about a performance to their students. How can this technology be best utilized and is the augmented feedback received by the student beneficial to their learning?

MakeMusic, Inc. (2005) makes the following claim in their advertising literature and on their website regarding the assessment portion of the *SmartMusic* program: "Assessment keeps students playing right by (1) On screen assessment shows students what they did right and wrong

(2) Hear solo part as needed (3) See fingering chart just by clicking on any note (4) Accompaniments provide steady tempo and pitch reference (5) Students learn to listen." Although these statements are used extensively in their literature, we were unable to find any empirical studies regarding the assessment feature of their software. Several studies have been conducted in the past regarding the accompaniment portion of the software, including the early *Vivace* system, and they generally indicate that the program was both beneficial to learning and to students' musical growth (Glenn, 2000; Ouren, 1997; Snapp, 1997; Tseng, 1996) but none of these studies addressed the relatively new assessment feature.

Although applications which give feedback based on performance have been in use for some time, it should also be noted that none of these applications offer feedback on the expressive aspects of the musical performance (Parncutt & McPherson, 2002). Software that is capable of utilizing automatic analysis of acoustical cues determined by statistical models that simulate emotion judgments in a user-friendly format was recently developed in a research project at Uppsala University (Juslin, Friberg, Schoonderwaldt, & Karlsson, 2004) and its implication in music education is still being considered.

Self-evaluation skills are an important part of music students' training; however, a clear method for teaching these skills has not been established. As musicians progress, they seem to acquire this ability on their own and are more proficient at self-evaluation at higher levels of study. For example, a college musician generally requires much less guided practice than the beginning student. Computerized assessment is becoming an increasingly predominant feature of music software, yet the efficacy of utilizing these assessment modules has not been verified. Therefore, the purposes of this study were to: (a) determine whether junior high instrumentalists are able to self-evaluate their performances accurately when compared to a CAI program's feedback *(SmartMusic*) (b) determine whether junior high students self-evaluation skills were affected by using a CAI program (*SmartMusic*) which included playback of student performances, graphical representation of errors, and quantitative feedback by the computer; (c) determine if a relationship between self-evaluation skills and musical achievement exists; (d) evaluate the accuracy of the feedback being given by the CAI program *(SmartMusic*); (e) consider whether the use of computer-assisted instruction *(SmartMusic*) can affect musical achievement.

Method

Participants included junior high students, grades 7 through 9 (N=18), from intact classes. Students' experience ranges (attribute variable) were 1 year of study (n=6), 2 years of study (n=9), and 3 or more years of study (n=3). The students were members of two different band classes that were scheduled by grade level rather than ability. All students in the study had received prior training using the *SmartMusic* software consisting of in-class sessions utilizing the training videos provided (*Basic Controls* and *Assessment*) as well as teacher demonstration and in-class modeling. Most students also had past experience with the program including solo contests, and home experience commensurate with their musical experience. Students used the program for solo accompaniment, playing exams, and homework assignments throughout their years in this band program. Students were comfortable with the use of a computer, program controls, microphone placement and sound adjustment.

The *SmartMusic* program was selected as the computerized assessment component for this study. Music notation for the performance can be read and performed from the computer screen. Computer assessment is optional and presented in a color-coded system that is easily interpreted

by the students and a quantitative score (percentage of notes played correctly) is calculated for each performance. Recording capabilities are controlled by the performer. A demonstration of the music to be performed is available through the *Solo* control of the program, and is performed using a piano timbre (General MIDI) for all instruments. The metronome feature of the program, *Metronome On,* was used and included a one-bar count-off. The accompaniment feature of the program was not used.

Computer equipment consisted of a Pentium 4 - 2.8 gigahertz processor with 1 gigabyte of RAM, running Windows XP (2002). The microphone used for recording students was the standard instrumental microphone sold by *MakeMusic, Inc.* for use with the *SmartMusic* program, and was positioned according to the manufacturer's recommendations for each instrument type.

The music selected for the purpose of the study was a 12-bar excerpt from an intermediate-level étude, taken from *Barret Oboe Method.* The "Moderato" étude was designated as audition material for a Jr. High honor band. The excerpt was input into the *Finale 2005* program, with two versions necessary for corrections of range. Two *performance assessment* files (.FPA) were then created using the conversion feature of the *Finale 2005* program and saved on the test computer's hard drive; the appropriate file was selected later according to woodwind or brass family.

One monitor was present in the performance area during all phases of the study to assist students with the program and setup, store student performances, and ensure integrity of the treatment. Monitors consisted of the teacher/researcher, an adult assistant, or a senior high school band member.

A pretest-posttest repeated measures design was used in the study. Self-evaluation accuracy and musical performance achievement served as dependent variables in the study. Two independent variables for the study were: Computer-Augmented Feedback (CAF) and Review of Recording (Listen). Additional procedures common to all treatment groups and utilized during the pretest/posttest phases were Perform and Rate. Treatment groups for the study included the following procedure combinations:

Perform/Rate/Feedback
Perform/Listen/Rate
Perform/Listen/Rate/Feedback

A random sample (lottery) was used to assign students to one of the three treatment groups. Descriptions of each procedure utilized in the various control groups follow.

During the *Perform* section of all pretests, treatment, and posttests, students entered the designated performance area one at a time. Students were allowed to play anything of their choosing for a brief period, in order to warm up. The microphone was then placed in the appropriate location according to instrument type by the monitor, sound levels on the computer were checked by the monitor, and students tuned using the *SmartMusic* tuning module, which has a graphic interface that resembles a standard chromatic tuner. The appropriate performance assessment file (woodwind or brass) was located on the computer's hard drive and opened using the *SmartMusic* program's Custom File tab. The proper instrument was then selected as an option in the program, which accomplished transposition of the exercise. Performance tempo was predetermined by the performance assessment file, and was set to Quarter note = 90 mm. The *View Assessment* option was turned off during each student performance. The *Cursor On* and *Metronome Click On* options were selected for all students. Through prior experience with

the program, these options were believed to help minimize possible incorrect scoring of large sections of the étude due to early rhythmic errors in the performance. After all options were properly set and verified by the monitor, the student performed the selection while watching the computer screen. For every student, the *Record* feature was used and the resulting sound record of the performance was labeled and saved. This procedure was utilized during all phases of the study (pretest, posttest, and treatment), on all participants regardless of treatment group.

The *Review of Recording (Listen)* procedure utilized the computer recording generated during the student's performance and was played back using the computer. Students receiving this treatment procedure were allowed to immediately review the recording of their performance. During playback of the recording, the monitor adjusted the volume level to off on the accompaniment track (metronome click), so that students heard only their own performance. All assessment features of the program remained off during this activity. After the students reviewed the recording, it was labeled and stored electronically. Students were asked to turn away or leave the room while the computer assessment option was initiated and the score and screen shot of the performance saved. This procedure was utilized during the treatment phase of the study, and only on members of Group 2—Perform/Listen/Rate.

All students (Groups 1, 2, and 3) participated in the rating task, which consisted of the student assigning themselves a numerical score (0 to 100 percent) based on their self-assessment of their performance: *Self-assessment (Rate).* These scores were recorded and stored along with the sound files for each performance. Monitors in the room were careful not to give any indication of the accuracy of the students' assessments. This procedure was utilized during all phases of the study (pretest, posttest, and treatment) on all participants in every treatment group.

Computer-Augmented Feedback (CAF) procedures consisted of students' reviewing the computer assigned numerical assessment (shown as a percentage played correctly out of total number of notes) and a graphical representation of the music with wrong notes shown in red. These assessments were computer generated during a student's performance and their display controlled through the program option: *View Assessment.* The monitor did not comment on student performance or subsequent self-evaluation. Students spent time reviewing the graphical representation of their performance before finally viewing the associated computer generated score. This procedure was utilized during the treatment phase of the study, and only on members of the appropriate groups (Group 1—Perform/Rate/Feedback and Group 3—Perform/Rate/Listen/Feedback).

No control group was used during this study so that all students involved might receive instruction. Instead, an extended pretest (*Perform/Rate*) consisting of three trials each served as the control factor. Posttests (*Perform/Rate*) consisting of three trials each were completed by each participant after the treatment phase of the study. During both the pretest and posttest periods, students were not allowed to view the *Computer-Augmented Feedback.* Students not receiving feedback during the treatment phase (Group 2—Perform/Listen/Rate) received additional instruction time after the completion of the study. Recordings, computer assessments, and student assessments were saved for all performances.

The study took place over a 6-week period as part of regular class meetings, which included a week-long school holiday period. During the first week of the study, students were introduced and taught the étude in large ensemble meetings with the teacher/researcher. This included performance demonstrations by teacher/researcher, explanation of components such as rhythms and phrasing, and approximations of performances such as clapping, counting, singing, fingering and articulating. Music was divided into phrases for purposes of introductory instruction and introduced one phrase per day, at a rate appropriate for the skill level of the students. Due to the

difficulty level of the etude, students needed guidance through the piece before attempting to utilize the program. Most students in this study would have been unable to successfully sight-read the piece. They were given copies of the music and asked to practice the music at home or at school, and to bring the music to school each day. The second week of the study, students continued to work on the étude on their own and performed the étude as a group during warm up.

The third week of the study (pretest period) students first listened to the internal demonstration of the piece in a standard MIDI piano timbre, then recorded individual performances (*Perform)* of the music using the *SmartMusic* program. After their performance, students were asked to assign a self-evaluative numerical score to their performance (*Rate).* Each student performed three pretests during this week, with a minimum of 24 hours between performances. The recorded performances and computer assessments were saved for all students. These pretests also served as a control for later comparison. The school was closed for holiday break during the next week, so the treatment phase of the study began one week after completion of the pretests.

The treatment period took place during the two weeks immediately after the holiday break. All students used the computerized assessment program to practice the étude a minimum of 3 times per week (once during each class session, with a minimum of 24 hours between attempts). Students in Group 1 (*Perform/Rate/Feedback*) performed the étude, assigned a self-evaluative numerical score (quantitative score on a scale of 0-100) and then received feedback (computer graphical assessment and quantitative score) for comparison to their own evaluation. Students in Group 2 (*Perform/Listen/Rate*) performed, listened to their recorded performance and assigned a self-evaluative numerical score. These students did not receive the feedback (computer graphical assessment and quantitative score). In Group 3 (*Perform/Listen/Rate/Feedback*), students performed, listened to their recorded performance, assigned a self-evaluative numerical score, and received feedback (computer graphical assessment and quantitative score) for comparison to their own evaluation.

Posttests for each student were completed in the final week of the study (posttest period) with each student again attempting the excerpt three times. Like the pretests, students performed the excerpt and evaluated their performance by assigning a quantitative score. During this period, they did not review the computer assessment or receive any type of feedback about their performance. CAF scores and recordings of the performances were saved for comparison to student self-evaluation.

During all phases of the study, a monitor was present in the computer room to assist students with the program and to control the treatment. Students were allowed to restart once per session, at the discretion of the monitor. Students were also allowed to practice the étude on their own during the study, but did not have access to the computer program.

This study asked the following question: (1) Is the feedback (CAF score) given to musicians by the computer-augmented instruction (*SmartMusic*) program reliable as compared to expert judges? (2) Are junior high instrumentalists able to self-evaluate their own performances accurately? (3) Are the self-evaluation skills of junior high instrumentalists affected by using the *SmartMusic* program which includes a combination of playback of student performances, graphical representation of errors, and quantitative feedback by the computer? (4) Does a relationship between self-evaluation accuracy and musical achievement exist? (5) Does the use of *SmartMusic* over a set period of time increase musical achievement?

Results

Self-evaluation accuracy and musical performance achievement were dependent variables in the study. Self-evaluation accuracy was determined by comparing students' self-evaluations with computer-augmented feedback (CAF) evaluations. A Self-Evaluation Accuracy score (SEA score) was calculated by subtracting the CAF score from the student's self-evaluation score (Student score - CAF score = SEA score). Musical performance achievement was measured using the CAF score only.

(1) Is the feedback (CAF score) given to musicians by the computer-augmented instruction (*SmartMusic*) program reliable as compared to expert judges?

Two band directors separately reviewed a randomly selected sample of student recordings (20% of 216 recordings) and scored those performances in regard to rhythm and pitch errors. Judges were asked to mark any note in the performance that they felt was an error (rhythm or pitch only), but to ignore musicality factors (phrasing, dynamics) when scoring the exercise, in order to keep measurements equal to that of the CAI program. Judges were also instructed to review the recording as many times as they felt necessary to give an accurate assessment.

Reliability was calculated for each of the judges by counting agreements/disagreements with the CAF scoring for each note in the exercise. Instances of agreements were calculated when either the CAI program and judge both scored the note as "correct" or both scored it as "incorrect." Disagreements occurred whenever a judge and the CAI program scored the note differently. Mean reliability for the two judges' scores and the CAF scoring equals 74.5%. Spearman correlations between CAF scoring and judges' scoring also indicate a strong positive relationship between the CAF scoring and judges' scoring, Judge 1, $r=.9181$, $p<.0001$; Judge 2, $r=.8136$, $p<.0001$.

(2) Are junior high instrumentalists able to self-evaluate their own performances accurately?

Pretest CAF scores were subtracted from students' self-evaluations to yield a normalized score without regard to the achievement level for all pretest attempts (Student score - CAF score = SEA score). By only examining the student's degree of error in self-evaluation score from the CAF score, performance achievement on the étude was removed from the calculation. Positive numbers in the SEA score indicate students who overestimated their own ability; negative numbers indicate underestimation. A SEA score of zero is the ideal and indicates the student's subjective score matched the computer's objective score (CAF). The mean pretest SEA score was 3.48, SD=18.256. As illustrated in the distribution chart (Figure 1), the largest single group of scores is 15-20 points above the CAF score, and the second largest group falls 5-10 points below the CAF score. This indicates that, although students' scores varied greatly, the majority of students tended to score their performance 10-20% differently than the CAF score. This degree of error in scoring is the equivalent of one to two letter grades in a standard grading system.

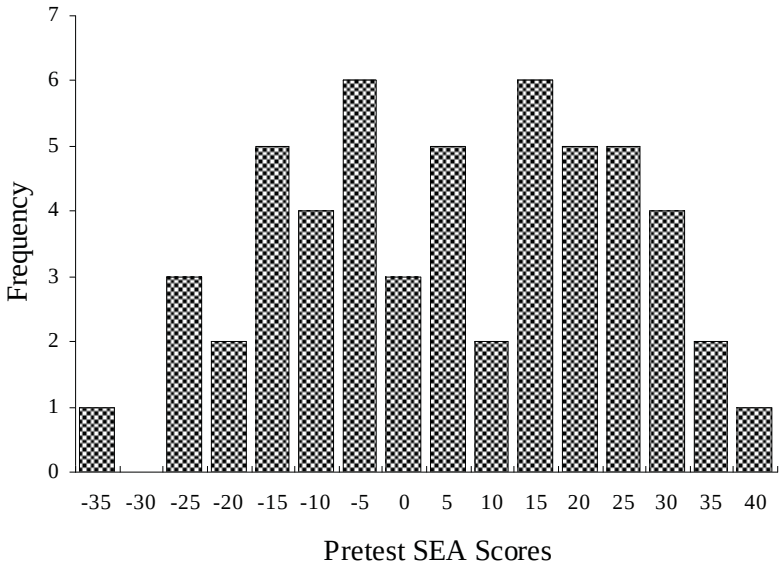

Figure 1. Distribution of Pretest Self-Evaluation Accuracy (SEA) Scores

(3) Are the self-evaluation skills of junior high instrumentalists affected by using the *SmartMusic* program which includes a combination of playback of student performances, graphical representation of errors, and quantitative feedback by the computer? If so, which combination has the greatest effect?

Mean pretest and posttest SEA scores were calculated for each subject and were then compared by use of a *t*-Test Paired Two Sample for Means; the overall results were not statistically significant. Although there were no statistical differences between groups, there was a trend for improvement of SEA scores in all groups (15 of 18 subjects). Because of the small sample size, a graphical trend analysis was also performed for each group (Figure 2). While these results are not statistically significant, Group 1 and Group 3 both had the greatest gains in student self-assessment ability (Figure 3).

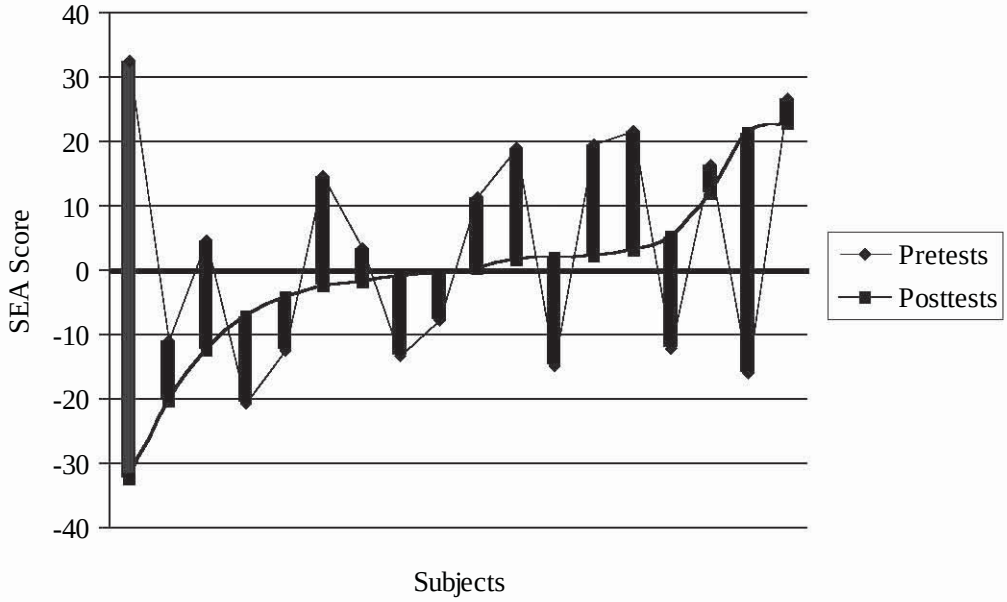

Figure 2. Mean Pretest vs.. Posttest Self-Evaluation Ability (SEA Score) by Subject

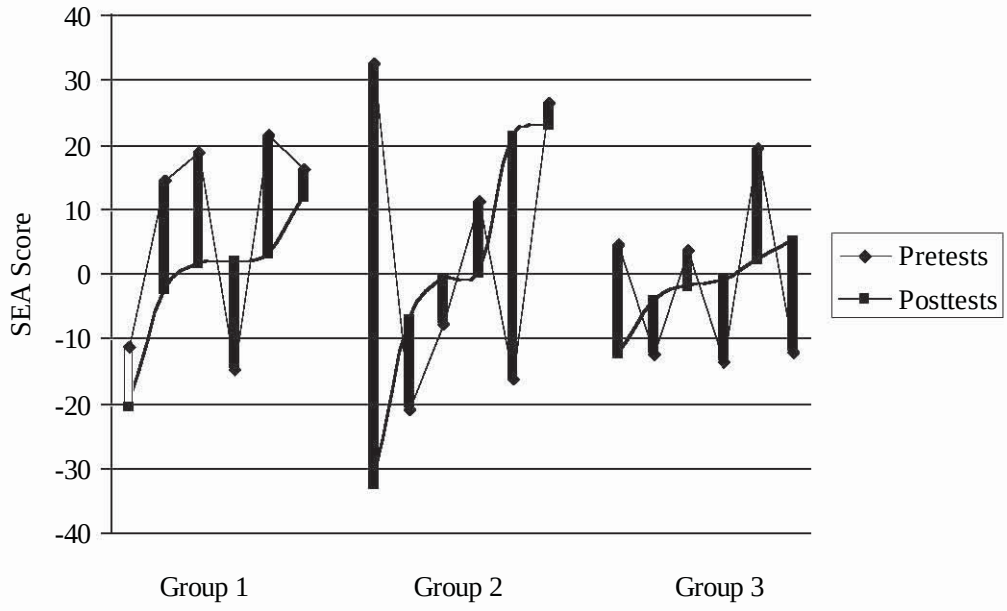

Figure 3. Mean Pretest vs.. Posttest Self-Evaluation Ability (SEA Score) by Group

(4) Does a relationship between self-evaluation accuracy and musical achievement exist?

Pearson correlation of students' SEA scores and the associated CAF scores, r=-.29, p=.0001, suggests a slightly negative relationship between students ability to self-evaluate and musical achievement. This correlation indicates that CAF and SEA scores generally moved in a somewhat converse direction. Of the 162 pairs analyzed, CAF scores ranged from 16 to 96 with a mean score of 64, SD=18.5. SEA scores ranged from -45 to 40 with a mean score of -1.11, SD=16.17. Matched data pairs of SEA scores and CAF scores for all attempts (pretest, posttest, and treatment phases) were also plotted (Figure 4). The scatter plot illustrates that while students with a higher achievement level do have slightly better assessment ability, students at all levels vary greatly in their ability to accurately score themselves. Students at higher achievement levels tend to judge themselves more critically than those performing at lower achievement levels. Students at lower achievement levels tended to overrate their own performance. This is illustrated in the scatter plot by the general downward shift in SEA scores as the CAF score increases.

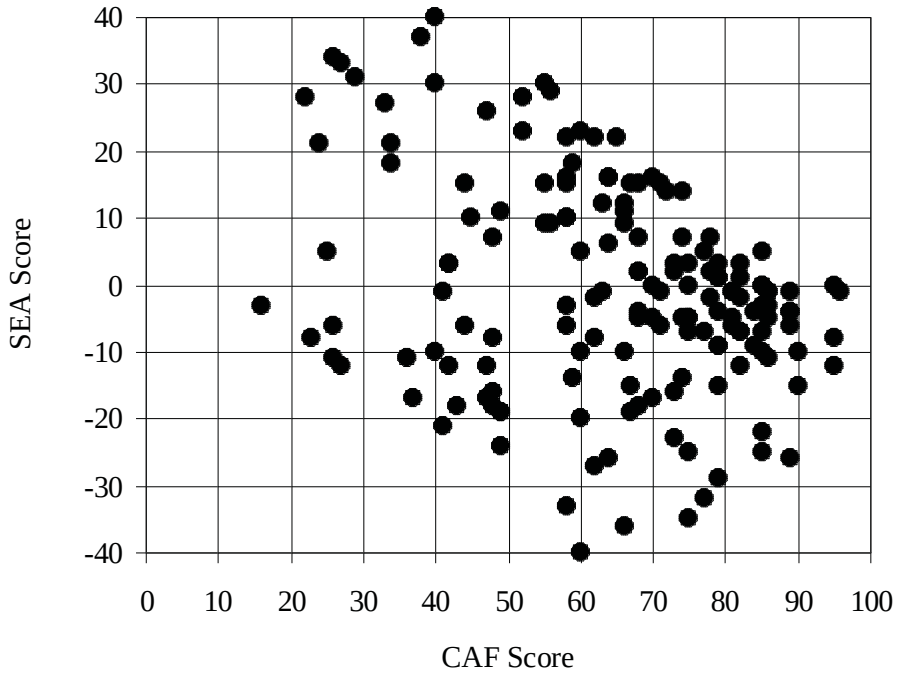

Figure 4. Self-Evaluation Ability Scores vs. CAF Musical Achievement

(5) Does the use of *SmartMusic* over a set period of time increase musical achievement?

Pretest and posttest scores were analyzed to determine if there was an effect on musical achievement. Comparison of the pretest and posttest CAF scores for each student showed that there was an improvement in musical achievement for all groups during the study, M=17.11, t=5.56, p<.0001. Paired t-Test analyses of the pretest/posttest CAF scores for each group are as

follows: Group 1, not statistically significant; Group 2, M=15.83, t=3.35, p <.0201; Group 3, M=26.00, t=4.64, p <.0056. Fifteen of the sixteen students involved in the study showed an increased average score in musical achievement from pretest to posttest (Figure 5). Although an increase in achievement was to be expected due to repeated practice, these results do indicate a significant increase in student musical achievement (Figure 6).

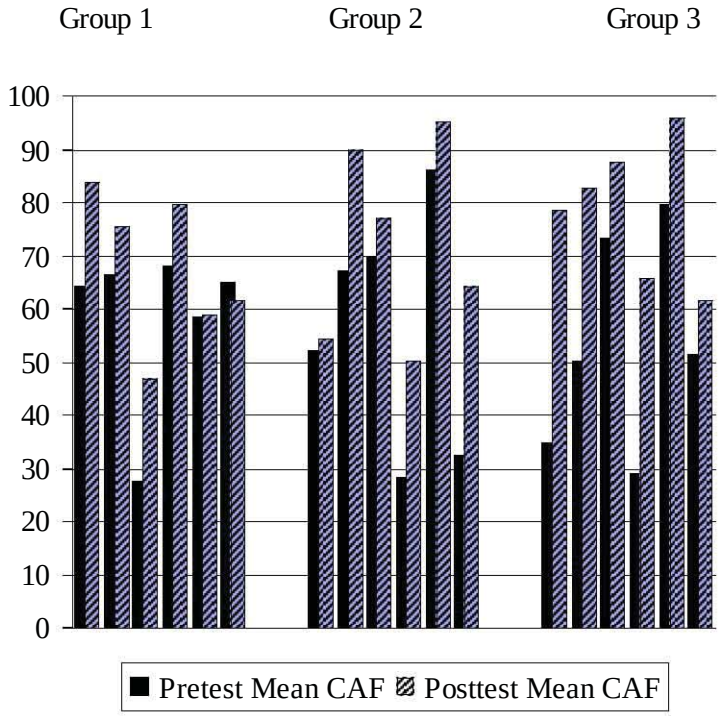

Figure 5. Pretest vs.. Posttest Musical Achievement Comparison

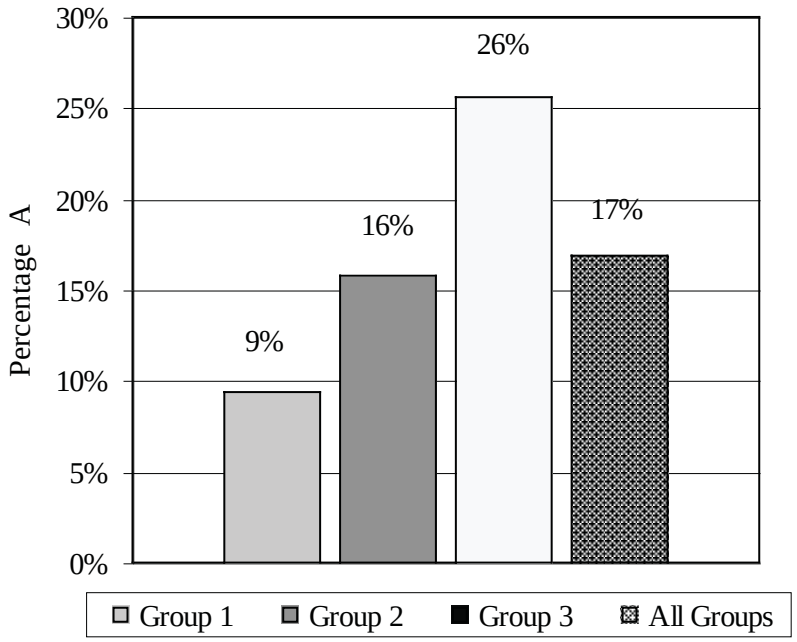

Figure 6. Changes in Musical Achievement Scores by Group

Reliability of the CAF *(SmartMusic)* and overall gains in musical achievement during the study were significant; however, because of the number of participants in this study, some statistical analyses of the data did not yield significant results. The initial pool of students included an additional five subjects; however, they did not complete the study due to injury or illness during the course of the study and any data for those students were discarded.

Limitations of this study are that there are no measures of musicality or of tone quality by the CAI program used, as it only determines accuracy of pitch and rhythm. Additionally, the sample size is too small to generalize results to a larger population. Concerns over reliability of musical achievement tests because of variables jeopardizing the internal validity also must be considered in the analysis of the data for musical achievement. These variables included the performance history during the study and the process of testing and treatment during the study. Nevertheless, conclusions may be drawn from inspection of subjects' and groups' changes over the course of the study. Pretest and posttest scores were analyzed to determine if there was an effect on musical achievement (quantitative analysis), although the internal validity of this portion of the study is known to be jeopardized. As the computer program only assessed pitch and rhythm, these were the only factors considered in the assessment of musical achievement. However, the internal validity of this study is known to be confounded by history and testing. Because students were also working on the selection outside of the treatment room, and because of the repetitive nature of the testing and treatment itself, increases in musical achievement should not be considered to be solely the result of the treatment.

There is a strong correlation between independent judges' scoring of the performance recordings and the CAF scores (.866) and also positive reliability of the scoring (74.5%). These findings will serve as a validity check for other data which relies on the CAF scores as

measurements of musical achievement, comparison to student self-assessments, and calculation of SEA scores. It was interesting to note that the correlation between quantitative scores is much higher than the reliability test. In examples that had very few mistakes, judges' scoring and CAF scoring were extremely close. As the number of performance errors increased, the quantitative scores remained very close but the *SmartMusic* program and judges disagreed on the actual placement of the error. Human judges seemed to count entire measures wrong when they heard multiple mistakes, whereas the CAF scoring was able to discriminate single correct notes.

Timing errors were often detected by the *SmartMusic* program and not by the expert teacher judges. Among the total disagreements recorded, judges marked 21% that the computer did not, and the computer marked 69% that the judges did not. This indicated that expert teacher judges may not be able to detect the very slight errors in rhythm or pitch that the *SmartMusic* program finds, or that the expert teachers are able to make accommodations for single note errors and the resulting juxtaposition of following rhythms. While these discrepancies exist, the accuracy of the CAF scoring seemed to be acceptable and the quantitative score assigned was very close to a judge's interpretation.

Analysis of pretests of all students' abilities to self-evaluate their performances in this study seem to agree with earlier studies which show that junior high musicians are unable to accurately assess their own performance. Students tested in this study generally scored themselves either 10 points lower or 20 points higher than their actual score. The range of error was -36 to +30, but there were only 7 scores out of the 54 pretest samples that were within a plus or minus 5 point range of the computer assessment.

While not statistically significant, a difference does exist between pretest and posttest self-evaluation skills. From this, we may conclude that student's ability to self-evaluate their own performances may be improved with practice. Utilization of computer-assisted instruction that includes feedback and recording abilities could be an effective method for practicing this skill. Exact combinations of features utilized in the computer-assisted instruction program seem not to be critical, but the treatment group that allowed a review of the recording before receiving feedback seemed to be more effective than the other methods utilized in both musical achievement gain and self-evaluation skills. Programs which offer this feature would be preferred over those that do not. A review of the recording alone, without feedback, had the least impact on the development of self-evaluation skills but did yield gains in musical achievement comparable to the overall study mean gain in musical achievement.

The data also suggest that a slightly negative correlation between musical achievement and self-evaluation skills exists, although also not statistically significant. Most students who performed very poorly had inflated self-evaluations of their performance. When students had CAF scores that were extremely low, they tended to miss that number in their own evaluations by rating themselves much higher. However, as student achievement on the étude increased, students then tended to rate themselves lower than the actual score. There was a tendency to neither be extremely critical nor extremely complimentary of one's performance. These students rated themselves higher than the CAF score when they did poorly and lower than the CAF scores when they performed well.

Students in this study seem to enjoy working with this program, although they were all extremely hesitant to evaluate themselves during the first few trials. Regardless of the quality of their performance, students seemed uncomfortable with the process of determining a score for themselves. As the study progressed, this uneasiness seemed to dissipate and students were able to complete the task of assigning a score much more quickly after the performance. From this, the conclusion seems to be that students may need more opportunities to evaluate themselves and

to verbalize the evaluations. Having the opportunity to evaluate and then compare the evaluation for validation seemed to be helpful in alleviating their self-consciousness with the task as well as improving self-evaluation skills. As students became more comfortable with the process, they also spent more time looking at the feedback and some even challenged items marked either correctly or incorrectly by the computer.

Additional research is needed to examine more closely the relationship between musical achievement and self-evaluation skills, the long- and short-term effects on self-evaluation skills when utilizing CAI programs, and ability to self-evaluate performance on non-repetitive tasks (sight-reading). It would also be useful to examine the effects on musical achievement of various practice techniques: multiple trials with no inter-trial delay or augmented feedback versus fewer trials accompanied by an extended inter-trial delay period consisting of a review of each performance recording and feedback. Research into human interpretation of errors and how we respond to multiple errors could also be beneficial in understanding the accuracy of CAF scoring.

The results of this study may have some practical implications for students and teachers. Computer-assisted instruction in music has developed from early software programs that consisted of basic tutorials and recording applications into sophisticated programs with the ability to assess performances in real-time. Use of these computer-assisted instruction programs seems to be beneficial to both musical achievement and self-evaluation skills of young musicians, and can provide reasonably accurate assessments when an instructor is not available. Receiving feedback after a performance, during the inter-trial delay period, and participating in self-evaluation during the inter-trial delay period both seem to promote an increased awareness of one's performance and foster an increase in musical achievement. Reviewing a recording of one's own performance can be beneficial to musical achievement, but needs to be accompanied by meaningful feedback in order to foster good self-evaluation skills. Finally, self-evaluation is a skill that students need to practice in order to feel comfortable and gain proficiency.

References

Anderson, J. (1981). Effects of tape-recorded aural models on sight-reading and performance skills. *Journal of Research in Music Education, 29*, 23-30.

Atchison, 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*. Unpublished Doctoral Dissertation, University of Iowa.

Barrett, A. (n.d.). *Oboe Complete Method*. New York, NY: Boosey & Hawkes, Inc.

Bergee, M. J., & Cecconi-Roberts, L. (2002). Effects of small-group peer interaction on self-evaluation of music performance. *Journal of Research in Music Education, 50,* 256-268.

Brick, J. (1984). An exploratory study of the effects of a self-instructional programme utilizing the Pitch Master on pitch discrimination and pitch accuracy in performance of young trombonists. *Psychology of Music, 12*(2), 119-125.

Byo, J. L. (1998). *Assessment of student performance in the ensemble setting*. Paper presented at the Graduate Music Education Research Seminar, University of Arizona, November 1998.

Byo, J. L. (2001). Designing substantive playing tests – a model. *Music Educators Journal, 88*(2), 39-44.

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*,

Colwell, R. (1970). *The evaluation of music teaching and learning*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

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: A project of the Music Educators National Conference* (pp. 1128-1158). New York, NY: Oxford University Press, Inc.

Colwell, R., & Goolsby, T. W. (1992). *The teaching of instrumental music* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Dickey, M. (1991). A comparison of verbal instruction and nonverbal teacher-student modeling in instrumental ensembles. *Journal of Research in Music Education, 39*, 132-142.

Dickey, M. (1992). A review of research on modeling in music teaching and learning. *Bulletin of the Council for Research in Music Education, 113*, 27-40.

Finale Performance Assessment (Version 2.0.0.60) [Computer software] (2005). Eden Prairie, MN: MakeMusic!, Inc.

Garafalo, R. (1995). *Instructional designs for middle/junior high school band* (Vol. II). Fort Lauderdale, FL: Meredith Music Publications.

Glenn, S. (2000). *The effects of a situated approach to musical performance education on student achievement: Practicing with an artificially intelligent computer accompanist*. Unpublished Dissertation, University of Georgia, Athens, Georgia.

Gonzo, C., & Forsythe, J. (1976). Developing and using videotapes to teach rehearsal technique and principles. *Journal of Research in Music Education,* 24, 32-41.

Henley, P. (2001). Effects of modeling and tempo patterns as practice techniques on the performance of high school instrumentalists. *Journal of Research in Music Education, 49*(2), 169-180.

Hewitt, M. P. (2001). The effects of modeling, self-evaluation, and self-listening 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.

Hewitt, M. P. (2005). Self-evaluation accuracy among high school and middle school instrumentalists. *Journal of Research in Music Education, 53*(2), 148-161.

Hickey, M., & Webster, P. (1999). MIDI-based adaptation and continued validation of the *Measures of Creative Thinking in Music* (MCTM). *Bulletin of the Council for Research in Music Education, 142*, 93-94.

Hickey, M., & Webster, P. R. (2001). Creative thinking in music. *Music Educators Journal, 88*(1), 19.

Hoffmann, J. A. (1991). Computer-aided collaborative music instruction. *Harvard Educational Review, 61*(3), 270.

Juslin, P., Friberg, A., Schoonderwaldt, E., & Karlsson, J. (2004). Feedback-learning of musical expressivity. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 247-270). New York: Oxford University press.

Kohut, D. L. (1985). *Musical performance: Learning theory and pedagogy*. Englewood Cliffs, NU: Prentice-Hall.

Linklater. (1997). Effects of audio- and videotape models on performance achievement of beginning clarinetists. *Journal of Research in Music Education, 45*, 402-414.

Magill, R. A. (2004). *Motor learning and control: Concepts and applications* (7th ed.). New York: McGraw-Hill.

MakeMusic!, Inc. (2005). *SmartMusic*. Retrieved September 10, 2005, from http://www.smartmusic.com/benefits.aspx

Meeuwsen, H., Flohr, J., & Fink, R. (1998). *Computerized assessment of synchronization and the imitation and timing of rhythm patterns*. Paper presented at the Fifth International Conference on Technological Directions in Music Education, San Antonio, TX.

Ouren, R. W. (1997). *The influence of the VIVACE accompaniment technology on selected middle school instrumental students*. Unpublished Dissertation, University of Minnesota.

Parncutt, R., & McPherson, G. (2002). *The science & psychology of music performance: Creative strategies for teaching and learning*. Oxford, NY: Oxford University Press.

Peters, G. D. (1993). Computer-based music skills assessment project: A portal to artistic innovation. *Bulletin of the Council for Research in Music Education, 117*, 38-45.

Puopolo, V. (1971). The development and experimental application of self-instructional practice materials for beginning instrumentalists. *Journal of Research in Music Education, 19*, 342-349.

Rosenthal, R. K. (1984). The relative effects of guided model, model only, guide only, and practice only treatments on the accuracy of advanced instrumentalists' musical performance. *Journal of Research in Music Education, 32*(4), 265-273.

Rosenthal, R. K., Wilson, M., Evans, M., & Greenwalt, L. (1988). Effects of different practice conditions on advanced instrumentalists' performance accuracy. *Journal of Research in Music Education, 36*(4), 250-257.

Sebald, D. (2000). *Developing automated student assessment instruments for the world wide web*. Paper presented at the Seventh International Technological Directions in Music Learning Conference, San Antonio, TX.

SmartMusic (Version 9.0) [Computer software] (2005). Eden Prairie, MN: MakeMusic!, Inc.

Snapp, D. R. (1997). *The used and effectiveness of the Vivace Intelligent Accompanist system in K-12 instrumental music programs*. University of Northern Colorado, Colorado.

Swinnen, S. P. (1990). Interpolated activities during the knowledge-of-results delay and post-knowledge-of-results interval; Effects on performance and learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*(4), 706-716.

Tseng, S.-M. A. (1996). *Solo accompaniments in instrumental music education: The impact of the computer-controlled Vivace on flute student practice*. Unpublished Dissertation, University of Illinois at Urbana-Champaign.

Venn, M. L. (1990). *An investigation of the applicability of recent advances in computer technology to the development of a computer-based, random access audio test of common criterion-referenced objectives in elementary music*. Unpublished Doctoral Dissertation, University of Illinois at Urbana-Champaign.

Young, C. R., & Regenberg, D. (2001). Multitrack cassette recorders: A technology for assessment. In *Spotlight on assessment in music education* (pp. 92-93). Reston, VA: MENC.

Zurcher, W. (1975). The effect of model-supportive practice on beginning brass instrumentalists. In C. Madsen, R. D. Greer & C. H. Madsen (Eds.), *Research in music behavior: Modifying music behavior in the classroom* (pp. 131-139). New York: Teachers College Press.

Edited by
Mary Ellen Cavitt, Texas State University—San Marcos

Relationships Between Attitudes Toward Classroom Singing Activities and Assessed Singing Skill Among Elementary Education Majors

Charlotte P. Mizener
Lamar University

Singing is an activity that is common the world over among people of all ages, all cultures, and all times. The creators of the National Standards for Music listed singing as the first of the Content Standards because they considered it the most important of all the musical activities (Consortium of National Arts Education Associations, 1994). Singing routinely takes place among children in the grade-level elementary school classroom.

In order for grade-level teachers to make use of music as a teaching tool, the training of elementary grade-level teachers usually involves at least one course in music fundamentals and music teaching methods. Students enter such courses with a wide range of musical backgrounds, skills, and attitudes. In most music methods courses, students are expected to sing individually while teaching songs and making other teaching presentations. Assignments involving singing often cause a great deal of anxiety among the students, and some may even drop the course because the prospect of solo singing is very distressing. Upon assuming their professional positions, these students often are expected to use music to reinforce academic studies or to use it as a "break" from other parts of the daily routine. When teachers are not comfortable singing, they may not use music to their best advantage in the classroom.

Several studies have investigated factors related to such areas as attitude toward teaching music (Kvet & Watkins, 1993), attitude toward importance of musical skills (Saunders & Baker, 1991), pitch-matching skills (McCoy, 1997), and grades in music theory (Harrison, 1996). Few studies have looked directly at development of attitude toward singing among adults, specifically elementary education majors.

The purpose of this study was to examine possible relationships between attitude toward classroom singing activities among elementary education majors and their actual singing skills. The study also investigated past musical experiences in the home and at school and their relationships to attitudes toward singing.

Method

Interdisciplinary studies students ($N = 20$) enrolled in elementary music methods classes in a medium-sized state university comprised the participants in this study. The participants were all female and came from two intact sections of methods classes taught by the researcher. They completed a questionnaire requesting information regarding past singing experiences within their families, singing participation in their school years, attitudes toward classroom singing activities, and their potential use of singing in their future classrooms. Most responses to the questionnaire items were marked on a five-point Likert-type scale. For these items, a response of four or five was considered more positive or more frequent; a response of one or two, less positive or less frequent. A response of three was considered neutral or average. A few items required a response of "Yes," "No," or "I don't know." Several other items requested completion of statements regarding singing preferences in childhood, and what other musical activities they might use in their future classrooms.

In addition, the researcher recorded the participants singing familiar songs and assessed their singing skills according to a rating scale developed by the researcher. The participants were recorded in the researcher's office during a regular methods class period. All recordings were made on a Sony audiocassette tape recorder. Each subject first sang "Jingle Bells" beginning on a self-chosen pitch. The subject hummed the first pitch before singing the song so the pitch could be identified in the analysis of singing skill. The subject then sang "Yankee Doodle" beginning on E4 (where middle C is C4). This beginning pitch was chosen because it allowed the song to fall in a range that would be comfortable for untrained singers (B3 to A4). The subject again hummed the initial pitch before singing. If she did not match it right away, the researcher worked with her briefly until she did match the given pitch.

Results

Data analysis included the calculation of means, frequencies, and percentages for some of the data and the use of crosstabulations for determining relationships among variables. A researcher-designed singing skill rating instrument was used in the analysis of the tape recordings.

Questionnaire Responses

Participants in this study generally had very positive attitudes toward singing and were moderately confident of their singing abilities, as shown in Table 1. When asked if they liked to sing, a large majority of participants (80%) indicated a response of 4 or 5, and 55% of the participants rated themselves at least average in singing skill. Table 2 shows that most participants had generally positive experiences with singing at home. In Table 3a, the data indicate that singing experiences at school were also positive. Most participants (80%) had participated in regular classes with a music specialist. Furthermore, all participants indicated that they anticipated using music in their classrooms in the future (see Table 3b).

An examination of the relationships among the variables revealed only a few significant associations. The only significant relationship between assessed singing skill and the questionnaire items was for the skill rating on "Jingle Bells" and responses to "I am comfortable singing alone (to children, to friends)" (X^2 [12, $N = 20$] = 22.37, $p < 0.034$, Cramer's $V = 0.61$). Participants with higher singing skill ratings were more comfortable singing alone.

Table 1
Frequencies and Percentages of Attitude Toward Singing and Opinion of Singing Skill Responses

	Questionnaire Item	Rating					
1.	Attitude Toward Singing ("Do you like to sing?")						
		Singing Attitude					
		3		4		5	
		n	%	*n*	%	*n*	%
		4	20	8	40	8	40
2.	Opinion of Singing Skill ("How would you rate your singing?")						
		Opinion of Singing Skill					
		1		2		3		4	
		n	%	*n*	%	*n*	%	*n*	%
		5	25	4	20	10	50	1	5

Table 2
Mean Responses to Questionnaire Items about Singing Experiences in the Family and the Community

	Questionnaire Item	Response
3.	When I hear songs on the radio, I enjoy singing along.	4.80
4.	I like to sing when I am by myself.	4.70
5.	I like to sing with a few friends.	3.60
6.	I am comfortable singing informally in a group (with friends, in the congregation at church, among the spectators at a sporting event).	3.65
7.	I think that I am as good a singer as most adults.	2.40
8.	I think that I sing as well as most people in my family.	3.55
9.	Members of my family sang for fun when I was growing up.	3.3
10.	Someone in my family sang to me when I was a child.	3.60
11.	I used to sing songs with some of my family when I was a child.	3.55
12.	When I was a child, I knew of several adults in my family who liked to sing and/or sang in a choir.	2.85
13.	Members of my family encouraged me to sing when I was growing up.	2.90

Table 3a
Mean Responses to Selected Questionnaire Items about Singing Attitudes and Experiences at School

	Questionnaire Item	Response *M*
14.	Singing (either with a music teacher or grade-level teacher) was an important activity to me in elementary school.	4.10
15.	Generally, singing in the elementary school was a pleasant experience for me as a young child.	4.20
16.	I am comfortable singing alone (to children, to friends).	3.65
17.	I am (would be) comfortable singing and teaching songs to children.	4.30
20.	Grade-level teachers should be able to sing, at least adequately.	3.45
21.	It is important for grade-level teachers to sing with and to their children.	4.35

Table 3b
Frequencies and Percentages of Responses to Selected Questionnaire Items about Singing Attitudes and Experiences at School

Questionnaire Item	Response					
19. Did you have regular music classes with a music specialist in elementary school?	Yes		No		Not sure	
	n	%	*n*	%	*n*	%
	16	80	2	10	2	10
24. When you become a grade-level teacher, do you see yourself using music in your classroom?	Yes		No			
	n	%	*n*	%		
	20	100	0	0		

Attitude toward singing ("Do you like to sing?") was also related to only one questionnaire item ("I think that I sing as well as most people in my family.") (X^2 [8, N = 20] = 16.05, < 0.042, Cramer's V = 0.63). It appeared that participants with a less positive attitude toward singing have a lower perception of their singing abilities compared to those of other family members.

The self-assessed singing rating ("How would you rate your singing?") had significant relationships to two items. In comparing the participants' own ratings of singing skills to responses to the item, "I think that I am as good a singer as most adults," it seemed that participants with a lower singing rating also tended to have lower perceptions of their singing

skills relative to those of other adults (X^2 [12, $N = 20$] = 25.97, $p < 0.011$, Cramer's $V = 0.66$). The item "Grade-level teachers should be able to sing, at least adequately" also had a significant association with self-assessed singing skill. Half of the participants thought that it was moderately important (a response of 3) for teachers to sing at least adequately (X^2 [9, $N = 20$] = 18.03, $p < 0.035$, Cramer's $V = 0.55$).

The majority of participants indicated that they were apt to use all of the classroom music activities listed in the questionnaire (see Table 4), and many included additional ones that they would be likely to use. A common suggestion was to use singing to accompany classroom routines (clean-up time) and transitions (moving from one place to another or changing from one activity to another). Another suggestion was to choreograph songs. The responses indicated that participants had a very positive attitude toward the use of music in the classroom.

Table 4
Frequencies and Percentages of Responses to Use of Music in the Classroom

Questionnaire Item		*Response*			
		Yes		No	
		n	%	*n*	%
1.	Teach songs or rhymes to help students remember academic information (like "The Alphabet Song," "Fifty Nifty United States," or the songs from "Schoolhouse Rock")	20	100	0	0
2.	Teach songs from or about the historical events that the class might	16	80	4	20
3.	Play recorded music from the historical periods that the class might study	15	75	5	25
4.	Teach songs from other countries that the class might study	15	75	5	25
5.	Play recorded music from other countries that the class might study	13	65	7	35
6.	Teach singing games and folk dances from other countries that the class might study	16	80	4	20
7.	Teach singing games and folk dances from American traditions	19	95	1	5
8.	Teach traditional songs that children should know, just for fun	18	90	2	10
9.	Teach holiday songs	19	95	1	5
10.	Play recorded music while the children draw pictures	19	95	1	5
11.	Allow children to play classroom instruments to accompany their singing	16	80	2	10
12.	Allow children to make up musical accompaniments for activities such as classroom plays	16	80	4	20
13.	Allow children to make up musical accompaniments for stories to reinforce reading lessons	15	75	5	25
14.	Allow children to explore classroom instruments or to make instruments to reinforce science lessons	15	75	5	25
15.	Allow children to make up tunes for poems or rhymes that they are reading and then perform the poem or rhyme	18	90	2	10

Assessment of Singing Skill

Overall, the participants demonstrated a relatively high degree of competence in singing as shown in Table 5, and all were able to match a given pitch for "Yankee Doodle." Their singing skill was assessed according to a revised version of a rating scale developed by the researcher for a previous study. Maintenance of tonality as determined by the beginning and ending pitches and by the accuracy of intervals are the primary criteria for singing skill in this rating scale (see Figure 1.) The beginning and ending pitches were identified by listening to the cassette recordings of the songs and comparing those pitches to pitches on a set of chromatic melody bells. The majority of participants (n = 15) started the song on a pitch within the minor third between C# 4 and E4. Other beginning pitches included Ab3, A3, B3, C4, and F4. Almost half of the participants (45%) scored a 6 on the rating scale for "Jingle Bells," and 60% of them scored a 6 on "Yankee Doodle." The rating of 6 indicated that the subject ended the song at a pitch no more than a half-step away from the beginning pitch and sang some noticeably inaccurate intervals, but she did not abruptly change tonality. In other words, the second phrase did not suddenly start at a different pitch. Instead, over the course of the song, the intervals were successively slightly out of tune until the song ended in a different tonality. Whereas all participants were able to match the beginning pitch for "Yankee Doodle," a few did not stay at the given pitch when they started the song.

Table 5
Frequencies and Percentages of Assessed Singing Skill for Two Songs

Song	Rating of Singing Skill							
1. Song 1, "Jingle Bells," self-chosen starting pitch	5		6		8		9	
	n	%	*n*	%	*n*	%	*n*	%
	5	25	9	45	4	20	2	10
2. Song 2, "Yankee Doodle," researcher-chosen starting pitch	5		6		8		9	
	n	%	*n*	%	*n*	%	*n*	%
	5	25	12	60	2	10	1	5

9. Begins and ends in same tonality, with no loss of tonality within the song and no noticeably inaccurate intervals.

8. Begins and ends in same tonality with no loss of tonality within the song but with some noticeably inaccurate intervals.

7. Begins and ends in same tonality but with loss of tonality within the song and some noticeably inaccurate intervals.

6. Ends in tonality different by no more than a half-step from beginning tonality, with some noticeably inaccurate intervals.

5. Ends in tonality different by more than a half-step from beginning tonality, with some noticeably inaccurate intervals.

4. Ends in tonality different from beginning tonality, with an abrupt shift in tonality within the song.

3. Begins and ends in same tonality but with little pitch variation around the tonal center.

2. Ends in tonality different from beginning tonality, or ends in spoken tones, with little pitch variation around the tonal center.

1. Has no clearly established tonal center and most intervals are inaccurate, or is chanted in spoken tones.

Figure 1. Singing Accuracy Rating Scale. This rating scale was first developed by the author in 1990 and has been revised to meet the needs of the current study.

Discussion

The pre-service grade-level teachers in this study showed a generally positive attitude toward music and singing through their responses to the questionnaire items. Similarly, their experiences with music in elementary school were quite positive overall. Their attitudes toward music appeared to be independent of their assessed singing skills. Analysis of the data yielded only one significant relationship between assessed singing skill and items on the questionnaire. All singers were able to match pitch, and most could stay in tune within a half-step of the starting pitch. It is possible that with this relatively small group of participants ($N = 20$) and a comparatively high level of singing skill, significant differences would not occur.

There were very few significant relationships between the other variables selected for analysis. A previous study (Mizener, 2005) also found that rating of singing skill was not associated with items regarding family singing activities or singing activities in elementary

school. The significant relationships that emerged in the present study were not unexpected. It is understandable that those who have stronger singing skills would be more comfortable singing alone. It is interesting that it was the singing skill rating on only one song that yielded a significant association with a questionnaire item. The mean rating of the song for which the participants used a self-chosen starting pitch ($M = 6.45$) was lower than that of the song with the researcher-given starting pitch ($M = 6.7$). The differences in singing ratings may have resulted in the single significant relationship.

Less positive attitudes toward singing and self-assessment of singing skill logically would be associated with less favorable comparisons to other singers. It is unclear why attitude toward singing was related to less favorable comparisons with adults in the family, and opinion of singing skill was related to adults in general. Perhaps a larger, more diverse group of participants would yield different results. Further research is warranted.

In addition, opinion of singing skill was also related to beliefs that grade-level teachers should be able to sing at least adequately. With over half of the participants rating themselves at a 3 or 4 in singing ability, it is again not surprising that they would think with moderate conviction that teachers should be able to sing.

It appears that the pre-service grade-level teachers in this study will be very likely to sing and use a wide variety of musical activities in their classrooms. Their singing skills do not seem to affect their willingness to plan for classroom musical activities. To extend the research, the participants will complete the questionnaire again at the end of the semester after they have participated in lessons in which singing, moving, and listening activities are integrated with other academic areas. Perhaps they will select more of the listed classroom music activities to include in their classroom routines. Results may then suggest directions for elementary methods class curriculum in terms of singing activities.

References

Consortium of National Arts Education Associations (1994). *National Standards for Arts Education*. Reston, VA: Music Educators National Conference.

Harrison, C. S. (1996). Relationship between grades in music theory for nonmusic majors and selected background variables. *Journal of Research in Music Education, 44* (4), 341-352.

Kvet, E. J. & Watkins, R. C. (1993). Success attributes in teaching music as perceived by elementary education majors. *Journal of Research in Music Education, 41* (1), 70-80.

McCoy, C. W. (1997). Factors relating to pitch-matching skills of elementary education majors. *Journal of Research in Music Education, 45* (3), 356-366.

Mizener, C. P. (2005, February). *Relationships between attitudes toward singing among elementary education majors and previous musical experiences.* Poster session presented at the annual meeting of the Texas Music Educators Association, San Antonio.

Saunders, T. C. and Baker, D. S. (1991). In-service classroom teachers' perceptions of useful music skills and understandings. *Journal of Research in Music Education, 39,* 248-261.