Texas Music Education Research 2002

Reports of Research in Music Education
Presented at the Annual Meeting of the Texas Music Educators Association
San Antonio, Texas, February, 2002

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

Edited by: Edited by Mary Ellen Cavitt,
The University of Texas at San Antonio

Published by the Texas Music Educators Association, Austin, Texas

CONTENTS

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

The Relationship of Goal and Reward Structure and the Meaning Non-select Choir Members Attach to Their Choral Experience

Susan Dill Bruenger
The University of Texas at San Antonio

As the director of a non-select high-school choir, I was frequently plagued with such questions as: Who are these students? Why are they here? Why are these particular students behaving in this way? What do these students expect to get out of this choir? It seemed certain that students would have a variety of reasons for how and why they participated in a non-select choir. It also seemed certain that choral programs would vary in the degree to which they are aligned with the self-perceptions of non-select choir students and meet their goals. My research interest became focused on determining why students in a specific non-select choir situation were behaving the way they were.

Even though there is a lack of research focusing on non-select choir students, a search of the literature did suggest that personal investment theory was an excellent tool with which to study behavior and the meaning behind it. Personal investment theory posits that the meaning students attach to a given activity in terms of sense of self, perceived options, and goals is an important determinant of the extent to which they will invest themselves in the activity (Maehr & Braskamp, 1986). This means that attempting to understand and influence motivation behavior or personal investment in the non-select choir would call for both the observation and the interpretation of the behavior patterns of students.

Maehr states that the specific context in which behavior occurs is critical because the context can encourage the students to hold certain meanings. He also suggests that motivation problems can occur when the environment is not compatible with the enduring meanings held by the participants.

Related research suggests that the kind of positive motivation behaviors choir directors would like to see in their students (on-task, high persistence, high intensity, with evidence of continuing motivation) would come from students who are task goal oriented (Meece, Blumenfeld, & Hoyle, 1988; Maehr & Nicholls, 1980); have a strong support system in which to establish their musical identity (Maehr & Braskamp, 1986); have a secure perception of control over their

destiny, which Maehr refers to as self-reliance (Maehr & Braskamp, 1986); and have a secure sense of competence and be strongly goal directed (Maehr & Braskamp, 1986). Maehr (1984) also suggests that students' perceived options should include choir as being available and appropriate for themselves.

Personal investment theory likewise suggests that the classroom context influences motivation behaviors such as the adoption of task goals, persistence, and continuing motivation (Ames, 1987; Maehr & Midgley, 1991; Maehr, 1992). Moreover, the following teacher controlled contextual elements encourage the development of task goals and have a positive effect on the meanings that students hold: a) teacher adaptation of a goal and reward structure that emphasizes the importance and meaning of various learning tasks and stresses individual effort and progress; b) teacher adaptation of a roles/power structure that allows for student autonomy which encourages persistence and continuing motivation; and c) high level teacher questioning and feedback patterns (Maehr & Midgley, 1991).

The importance of qualitative research is not in its generalization to a broad population of non-select choir students, but in its in-depth description of events in a particular situation. This paper will report on one individual teacher's goal and reward structure, which was one aspect of this particular context's effect on the students in this choir's personal investment (Bruenger, 1999).

In terms of personal investment theory, feedback in the form of evaluation is an important determinant of the meaning that a student attaches to a learning or performance situation. Rewards are an important source of feedback. The reward distribution system a teacher develops effects student interactions and goal adoption. The three goal structures discussed in the research literature are the competitive structure, the cooperative structure, and the individual structure (Ames, 1984; Nicholls, 1984). Rewards are distributed within these structures on the basis of the comparative performance between individuals, group accomplishment of a shared task, and the quality of the individual's task performance. Baden & Maehr (1986) contend that these structures have different effects on the behavioral and cognitive patterns of the students. Different structures prompt the development of divergent attributional patterns and strategy development (Ames, 1981; Diener & Dweck, 1978). These structures also have varied affects on the student's feelings toward school, self-concept and performance (Aronson et al., 1978; Johnson & Johnson, 1981). Moreover, these structures can affect the anxiety levels of students in varied ways (Hill, 1984).

Students' performances are valued by their position in relation to others in a competitive environment. Beneficially displaying one's skill and winning or "coming out on top" are deemed a priority. Also, the opportunity to attain one's goal or to receive the reward is reduced when other students are successful. Students in these competitive situations will not voluntarily pursue challenges that place their classroom status at risk.

Students tend to compete only when they feel competent. In a competitive environment, which limits the number of potential successes, a student who has experienced failure, and does not feel competent in the subject at hand may become anxious and try to avoid competing.

Some researchers have found that failure in competitive settings has more negative consequences for students' self-esteem than failure in noncompetitive settings (Ames & Felker, 1979). For example, children have recounted significantly lower levels of satisfaction following a failure in competitive situations and evaluate the poor performances of others as least satisfying when they involve a competitive loss. Ames (1984) has found that losing in competitive settings magnifies negative affect more than winning enhances positive affect.

While winning evokes self-aggrandizing motives, the findings are strong that failing in competitive structures elicits feelings of nondeservingness and dissatisfaction; and because competition engenders a situation of many losers and few winners, an esteem rating for this structure must necessarily be low. (Ames, 1984, pp.184-185)

A cooperative structure is characterized by students participating in groups where rewards or punishments are shared as a function of their combined performance. This goal/reward structure can be useful in integrating students of varying sociocultural backgrounds (Johnson et al., 1981). When a cooperative learning structure is well organized all students take responsibility for the completion of the task. The avoidance behavior of students with low sense of competence is reduced when accomplishment of a task, rather than social comparison determines success on task. "Through their participation in a cooperative task, low-achieving children increase their chances of actually learning the skills they may be lacking" (Baden & Maehr, 1986, p.299).

The third goal/reward structure is the individual. The participating students are independent of each other and everyone shares the same possibility of rewards that are distributed for individual improvement of performance. The difference between an individual structure and a competitive one is the emphasis on outcome and social comparison that is found in the competitive environment. Tasks are more likely to match a child's own competence level in the individualized setting than in either the cooperative or the competitive structure. Individual effort and progress toward task mastery are emphasized and performance comparisons are minimized (Baden & Maehr, 1986).

When both individualized contexts and competitive contexts have been evaluated, only individualized contexts have been shown to encourage student's utilization of their past performance in evaluating their present performance (Ames, 1984). This strategy has been shown to support the development of a task mastery orientation (Covington, 1984; Nicholls, 1978).

Nicholls asserts that the individual task involvement structure produces the most desirable outcomes, especially when trying to develop achievement in culturally diverse students (Nicholls, 1984). This structure also fosters challenge seeking (Nicholls, 1984) and continuing motivation (Maehr, 1976).

Ames (1984) views cooperation and competition as situational factors that influence a student's motivation. She maintains that "a competitive structure promotes an egoistic or social combative orientation, a cooperative structure elicits a moral orientation, and an individualistic structure evokes an achievement-mastery orientation"(Ames, 1984, p.189).

The sharing of goals, rewards, and effort as well as a positive interdependence among students are characteristics of the cooperative structures. Elements of the cooperative structure are seen in many non-select performing groups. The performance goal is shared by everyone. Rewards in the form of grades are shared for example, in Casey's (1994) study of non-select bands the majority of directors reported that they gave an A to non-select band students solely on the basis of effort as measured by attendance with no individual mastery required.

Ames (1984) contends that when a cooperative situation elicits norms for helping, aspects of a moral situation are present. She elaborates by stating that moral situations are essentially intent-oriented systems, behavior is judged by the student's willingness to put forth effort. Effort in a cooperative situation is not however, the individualistic conception of effort, in other words, working for one's own profit. Rather, the cooperative effort serves group goals and demonstrates social responsibility (Ames, 1984).

Ames (1984) also reports that students are often perceived as similar in ability and deserving of reward in a cooperative situation even when their performances vary in achievement level. The

group dynamic creates a perception of equality; and equality, more than equity, governs the distribution of rewards. "Thus whereas competitive situations foster perceptions of differences, cooperative structures contribute to perceptions of similarity" (Ames 1984, p. 196).

Ames cautions that such positive consequences, as heightened self-esteem that are attributed to cooperative structures may be a result of group outcome rather than the cooperative structure itself. In a statement that could be important to many non-select performance group situations Ames maintained that:

"We cannot predict self-evaluations solely by looking at individual achievements; children attend to group outcome information and the effect of negative outcomes deserve special attention in future research on cooperation." (Ames, 1984, p.187)

Ames describes the evaluation system in the moral domain as based on blame, rather than reward. Individuals are punished for not following social norms rather than rewarded for upholding them. Sometimes low achievers are targeted for blame when group members feel that they have failed to exercise proper commitment to the group. Ames research has determined that individual failure becomes most notable when the group fails. For example, Ames found that when group members were asked to distribute rewards, the low performer was given less reward when the group failed to accomplish its task than when it has been successful (Ames, 1981). The blame phenomena in cooperative structures only exist when the group fails, successful groups do not engender this reaction (Ames, 1984).

Method

The non-select choir examined in this study was a mixed chorus with a wide range of student behaviors. The school is in an upper-middle-class suburb of a large midwestern city. Over 15 years ago a federal court ordered the busing of inner city Black students to the wealthy county schools in this area. Even though white students are the majority ethnic group in the school, the ethnic makeup of this particular class was 50% Black and 50% white. This mixed chorus consisted mostly of freshmen students along with several others who had not auditioned for select groups. Some of the students were very serious choir students and had lofty musical aspirations, whereas others were in choir "just for fun."

The teacher in this class, Mrs. T (ethnicity: White), has a Masters degree in Music Education and 30 years of experience teaching, 25 of which have been in this district. Her choir program has a good reputation in the area with her select choirs typically receiving first place ratings at contest and various festivals. She has 65 students participating in her select choir who are, for the most part, models of achievement motivation. Even though this teacher (who is considered a strict disciplinarian by colleagues and students alike) manages to control her class, there were several students who constantly tested class rules and displayed only sporadic interest in singing.

Data Gathering Strategies

I used two video cameras and one audiocassette recorder when I observed, recorded, and described the direction, intensity, continuing motivation, and persistence behaviors of every student in the choir along with teacher practices. I initially observed and recorded student behaviors during each lesson in person, listened to audio-tapes of the lesson repeatedly in order to transcribe verbatim what occurred, and then viewed each lesson on video tape a minimum of two times to verify the accuracy of my observations. The following qualitative procedures were applied to assure the reliability and validity of my observations:

Stability of observations. Stability was addressed performing my observations over an entire semester. This was an important safeguard in this investigation because I was able to observe and document behavioral changes that might have been missed in a study of a shorter duration.

Consistency check. The consistency of my observations was addressed when I taught an external observer how to observe and record personal investment behaviors and context variables. Tapes of 10 % of the total observations were randomly selected for the external observer to score. This check resulted in an interobserver reliability of .93.

Follow up questions. I pursued answers to any questions generated by a behavior or comment made by a student or teacher during a particular class session. For example, I employed a follow up question when I asked Celeste and Bonita about their reluctance to use their music.

Feedback test. A feedback test involves soliciting feedback from a variety of people in order to identify theoretical threats, researcher biases and assumptions, and flaws in researcher logic or methods. I employed a feedback test when I questioned the students and this teacher about the low level questioning pattern I observed.

In order to determine the meaning each choir member attached to the experience relative to sense of self, perceived options, personal incentives, and other salient contextual factors perceived by the students, I interviewed students in small groups of two to three. I talked to them using open-ended questioning to encourage them to reveal the meanings that they attached to choir and give them the opportunity to bring up contextual factors that they found to be important to them.

Codes were used to reduce the data from interviews into an analysis of sense of self, goal orientation, perceived options, context effects and antecedents to meaning cited by the students. Codes were also used to reduce the data from observed behaviors into an analysis of direction, intensity, continuing motivation and persistence, along with an analysis of the context and the antecedents to meaning that affect the context. All of these codes were derived from personal investment theory

The following qualitative procedures were applied to assure the reliability and validity of interview data and the interpretation procedures:

Member checks. Member checks were completed to make sure that the information gathered to determine the meaning of choir to the students accurately fit the students' perception of the situation. Complete transcripts of their interviews and member check forms were presented to each student. The forms allowed for disagreement, agreement and for the opportunity to add something to their statements. Every student signed off on his or her form.

Verification questions. I verified all comments students made about events that occurred when I was not present. For example, I asked both Mrs. T and Celeste about the confrontation that took place when I was not present.

Coding check. The dependability of my coding scheme was addressed when I taught an external observer codes and then had each of us transcribe a random selection of 10% of the data. This resulted in an inter-rater reliability of .85.

Trustworthiness. This study is presented in a manner which allows the reader to follow the process of data collection and analysis from the transcripts through the coding process, to summary profiles, and ultimately, to my conclusions. It also enables the reader to decide whether or not to agree with my findings and conclusions.

Results

My observations are limited in this paper to a) a typical day's procedure, and b) the predominant features of the context in terms of this particular teacher's goal and reward structure. The observations are triangulated with student and teacher comments.

The open-ended interview style I used encouraged the students to respond freely and frankly to my questions and to describe events with their own emphasis. As a result, students gave a vivid description of the contextual features that were important to them. This was critical in my interpretation of how the many contextual features impacted upon the meaning of choir for these students.

Description of a Typical Class Session

The class met every day from 8:30 a.m. to 9:15 a.m. The educational goal most evident on a daily basis was the learning and perfection of choral literature for a "Classical" music concert on March 18th, in which each performing group at the school presented a five to ten minute performance, and a Spring "Pops" choral concert.

The class session always started with the teacher taking roll. It would frequently take Mrs. T five minutes to get roll taken and the students tended to socialize with each other to fill in this time. The students in this class were not particularly punctual and on a typical day one to three students were tardy. As long as students got into their seats before Mrs. T finished roll, she rarely commented on tardiness.

After taking roll, the teacher discussed pertinent class-related business, such as concert information. Besides preparing for the three concerts as described above, Mrs. T also took this class on two field trips. In April they went to a concert at the symphony hall and in May they sang at an amusement park. When the teacher finished with class business, warm-ups began, which lasted anywhere from 2 to 10 minutes.

Warm-ups were designed to cover breathing, resonance, range, and flexibility. The students willingly participated and, for the most part, enjoyed the warm-ups. Several students told me they thought the vocalises were beneficial. Mrs. T conducted her warm-ups as follows: first she demonstrated an exercise, then the class would imitate her, and finally they would together modulate up or down depending on the purpose of the exercise. Mrs. T limited the warm-up exercises to simple vocalises built on major scales and arpeggios. In my opinion, Mrs. T's greatest strength as a teacher was her beautiful voice and her ability to model good tone quality and diction and to reinforce and cajole her students until they imitated her. This group sang with a lovely sound due to her tireless efforts.

Warm-ups were followed either by interval drill and/or sight-reading in 5 of the 30 class sessions I observed. No classes after March 6th included sight reading. When the class did sight read, they used solfege and the teacher both sang and played their parts along with the students. When the teacher stopped playing, the students almost always stopped singing, Jean and Mic being the only exceptions. Persistence was extremely low for the class as a whole. The low persistence observed during formal sight singing exercises also occurred when the class was learning a new piece of literature. It seemed clear that they did not practice sight singing often enough to feel comfortable with the skill. Also, for most of the students, this class was their first exposure to sight singing as there had been no training in it at the junior high level. When they were asked to sight sing, the students grumbled, and discipline problems increased. The only students who sight sang well were students who played piano or some other instrument. Hence, musical literacy in the strictest sense was not a strength of this class.

Most of the time, warm-ups were directly followed by repertoire rehearsal. Music history/appreciation was sparsely addressed, although Mrs. T did show videos on Bach and Beethoven. She also talked about Brahms and invited the students to write and read to the class an extra-credit report on Brahms, which only Tona decided to do. Mrs. T's priority was developing her students' vocal technique and teaching the performance of choral literature.

Literature was learned mostly by rote with Mrs. T using a master/apprentice model in her teaching. With her lovely voice she consistently demonstrated the way she wanted her students to sing the music, and they imitated her.

Once the class began rehearsing literature, the demonstration/imitation format was followed until one or more of the students drifted off-task and either sat there not singing or began to hinder the instruction in some way. The teacher dealt with most off-task behavior with a verbal reprimand and most of the time it was sufficient to get the class back on track. However, sometimes the situation escalated when the student chose to disagree with the teacher's assessment of behavior. Disagreements and full-fledged confrontations were not rare. I witnessed 14 teacher/student disagreements and 16 teacher/student confrontations during the 29 class periods and 2 concerts I observed over the semester.

Typically, when a disagreement or confrontation occurred, Mrs. T stood her ground and if the student did not back down, he or she was asked to leave the classroom. In such instances, Mrs. T followed up with the disciplinary procedure recommended by the school and sent paperwork to the assistant principal. Two students, Ginny and Laverne, were permanently removed from class during the semester I observed due to confrontations with the teacher. Bonita, Celeste, and Jan were asked to leave class due to confrontations but were allowed to remain in the choir after consultation with administration. The only variations to the above routine happened: a) when Mrs. T showed videos; b) on two occasions, when she had the students stand in a circle and sing (the students really enjoyed this); c) when rehearsing "American Dance Party," which had choreography; and d) when the students rehearsed on risers in the auditorium the day of the classical music concert (March 18th) as well as for three days prior to the pops concert.

Mrs. T did not have an accompanist, so she conducted most of the class from behind the piano. Besides its obvious musical function, the piano unfortunately seemed also to act as a physical manifestation of a psychological barrier between Mrs. T and the students. This did not aid her in establishing a "we're in this together"-relationship with her students. She occasionally came out in front of the piano to demonstrate something and when she did, it always got the students' attention. As Jean and Jane stated, they thought Mrs. T's characteristic position at the piano emphasized the difference between teacher and student roles:

Jean: Really! She doesn't have a lot of contact with her class. I mean, it's just, "I'm up here." There's like a big barrier.
Jane: Yeah!
Jean: "I'm up here, and you're back there and we do not mix."
Jane: Exactly. [13/Jean/27]

Predominant Features of the Context in Terms of Personal Investment Theory

In this section I will describe my observations of one aspect of personal investment theory contextual features (see Table 1): goal and reward structure. My observations are substantiated by pertinent student and teacher interview excerpts.

Table 1
Observed Contextual Influences Frequencies

	Cooperative Goal Structure	Competitive Goal Structure	Individual Goal Structure	Group Task Reward	Comparative Reward	Individual Gains Reward
Total	9	5	0	15	0	0
Mean	1.0	1.3	0	1.4	0	0

Goal and Reward Structures

Table 1 displays total frequencies and means of my observations of teacher comments and behaviors indicating goal and reward structures. Goal and reward structures are created by the teacher and are best determined by analyzing which of the following the teacher emphasizes: doing better than other students (competitive goal structure), cooperating with other students toward a common goal (cooperative goal structure), or individual students mastering tasks (individual goal structure). Reward structures are also created by the teacher and can emphasize: a norm-based reward structure (comparative reward structure), a group accomplishment of a shared task reward structure (group accomplishment reward structure) or individual students mastering tasks reward structure (individual gains reward structure).

Mrs. T's instructional goal for this class was to teach them how to sing and perform western European choral music in the traditional manner and style. She concentrated her efforts on coercing the students to work together toward securing a group sound and group results. She most frequently addressed the students as a group not as a collection of individual singers. The following quote illustrates the emphasis she put on group rather than individual processes.

Mrs. T: Now, does everybody else want to sing, is that what you're here for? OK, then we need to do it together and do it right. Then it won't take up so much time.

Mrs. T never singled out individual students to praise or to correct their work. She never made comparisons between students. Nor did she, for example, ever suggest that a student serve as a vocal or behavioral model for others. She did not appoint section leaders and she didn't post or announce grades. Solo singing was almost absent. One time a solo appeared in a song they were sight singing and she asked for a volunteer to sing it rather than appoint someone herself.

Student and Teacher Comments on Goal and Reward Structure

The next four interview excerpts reveal students' observations of Mrs. T's emphasis on group processes and results. The first excerpt demonstrates Jane and Jean's chagrin when their

classroom performance was perceived as similar to others in the group and equally deserving of reward even when their performances varied in quality:

Mrs. B: Could you give me an example of what you mean by "things she beats into you?"

Jean: "Move your mouth"...everything. I mean, everything.

Jane: "Stand up straight."

Jean: And she goes about crazy it. And I understand it because a lot of people don't do it. There's got to be some reason, something she can do besides...

Jane: (interrupts)Yell at the whole class...

Jean: Yell at all of us. You know, I mean, I'm sure she's probably tried to think of something, but, to correct these people, but don't correct me in the mean time, I'm sorry, it's infuriating! [13/Jean/26]

In the next excerpt Bobbi reiterates Jean and Jane's observation that Mrs. T treated all of the students equally, even when they behaved differently.

Mrs. B: How about you Bobbi?

Bobbi: Um, I guess because of my experiences this year I'd have to say I won't be in choir again.

Mrs. B: Have you looked at A Cappella? Would it be more of the same?

Bobbi: I don't know if it would be the same but the teacher would be. And I really don't like how she treats everyone. Some people might deserve it but other people don't. [16/B/13]

Jean and Jane also described how the individuals in this class were not individually rewarded for their efforts to get along and do what they are supposed to do.

Jean: ... but she stereotypes all of us. I think that's what she kinda does in her mind and I sympathize with her. But she has to know who wants it and who doesn't.

Jane: Exactly. You can tell by the way people sing, by the way people… I mean I want to be there; I want to be there. I mean, you can tell by the way I sing. Maybe not by the way I talk to her, or the way I look at her, but by the way I sing.

Jean: She always does it; she always stands up, sings, does what Mrs. T asks of her and 20 times more, and got kicked out. She really does. I think a lot of us do, cause I will say, yes, I have tried so hard to please this woman. I think I've done a lot more than most people would, so I hope she sees that, I really do. [13/Jean/30]

The following excerpt is an example of the numerous comments the students made in their interviews on how Mrs. T punished them for not following her rules of behavior. I did not observe the teacher rewarding students for upholding these same rules of behavior and not one student mentioned being so rewarded.

Mrs. B: What you are saying is she doesn't tolerate mistakes you make in behavior, but she doesn't make you feel bad about musical mistakes.

Sheri: That's not bad.

Sharon: No, that's not bad. She just sits on the mistakes on things that we do wrong, but not the mistakes we make in singing. Like things that people do in class she just dwells…

Sheri: She just picks on them. Just like that thing with Jane, she just sits on them and then she'll never let it go until something else comes along.

Mrs. B: So you don't feel afraid of singing a wrong note. Because she's not going to pick on you for that.

Sheri: No. She'll just jump down your throat for something behavioral wise.

The first sentence of the next excerpt verifies my observation that the teacher did not single out individual students to compliment. This excerpt also gives Emily and Katherine's response to the single incident when a solo occurred in the music the class was sight singing and Mrs. T asked for a volunteer to try it. Emily volunteered to sing the solo and enjoyed the experience. She was also struck with the insight that the opportunity she took advantage of would also benefit other students in the class.

Emily: She never compliments anybody. She compliments us as a whole and um, I liked it when she had us do a solo; remember when she did that with that one spiritual?

Katherine: Yeah.

Emily: I thought that was a wonderful idea because it made their chance to shine. Because it seems to me, a lot of people don't get a chance to be in plays or get that feeling of "I did good" all the time. So if people got a chance to do a solo and say "Hey you did a good job today, would you do this solo today?" And that would make people think "Wow, she really did hear my voice so now I can stand up and show the rest of the class what she's talking about" and that's something where they could get a feeling of I did good, you know? [18/E/ 7]

When, in an interview with Mrs. T, I commented on the her lack of preferential treatment of her students, her initial response was, "They probably think I hate them all equally." She then elaborated as follows:

Mrs. T: I don't know. I'm glad that I don't. I guess I work at not trying to favor one over the other. Down through the years I've had…probably I guess it was 8 or 10 years ago? I had a student from the city, who…she didn't make A Cappella her first…you know, her sophomore year and then the next year she got into A Cappella, but she didn't make the show choir. And, uh, she accused me of being a racist and you know because she didn't make it.

And so she kinda like rallied her little friends around her and it was a real mess. Her mother worked for Congressman Smith on the north side. And I never did talk to the mother. Her mother never called. And never returned my calls and they got the ACLU or the NAACP—I honestly don't remember which one now. But, anyway they sent a representative to investigate me for discrimination. And it was horrible. I felt like, you know, that I was like under the glass every moment.

And I guess that's when I started examining myself, saying maybe I am doing this, maybe I don't know I'm doing it maybe. And it really wasn't that she just didn't sing as well. And she had a lousy attitude and that was the reason she didn't make it. And it really makes you stop and examine yourself. And after they sent their representative out, you know, and they saw how I auditioned kids and their rankings and stuff. You know this is why and of course its still subjective. Anytime you're talking about singing, you don't sing on pitch or in tune, or you didn't sight read that as well as somebody else.

And for a while I thought, well maybe I should get somebody else in to help with auditions. Of course that's a pain too because it's adding a whole new level to everything. But that was a really hard thing to go through.

Conclusions

Mrs. T's emphasis on cooperating with other students toward a common performance goal and her inclination to only give rewards (compliments) to the group as a whole for group

accomplishments of the shared task of performance lead me to classify the goal and reward structure observed in this context as cooperative goal and group reward structures.

As Ames (1984) suggests, one would expect to see the following in a class run with cooperative goal and group reward structures: a) the students will be perceived as similar in ability and deserving of reward even when their performances vary in quality; b) the students behavior is judged by their willingness to put forth effort, but not an individualistic concept of effort, rather a cooperative effort serving the group goals; and c) blame is a noticeable feature in that individuals are punished for not following social norms rather than rewarded for upholding them. All of the features Ames described were prominently featured in this particular context and were verified by the students in the excerpts above.

A positive aspect of the first feature that Ames described above (the students will be perceived as similar in ability and deserving of reward even when their performances vary in quality) was a classroom with a very high percentage of students who felt good about their ability to sing. I asked the students to rate themselves from 1 to 50, with 50 being the best student in the class and 1 being the least good student. The responses of the 31 students who finished the semester responded to this question are shown below:

Table 2
Student Self-Assessment of Singing Ability on a Scale of 1-50

Natalie		35	
*Nan	40		
Cissie		25 or 30	average 27.5
Margaret		35 or 40	average 37.5
Bonita		on both ends (1 or 50)	average 25.5
Regina		35	
Kim		45-48	average 46.5
Jan		30	
Ellen		29	
Celine		47	
Mary		41 or 42	average 41.5
Paula		39 or 40	average 39.5
Amelia		40's	average 45
Bobbi		30	
Celeste		45	
*Crystal		40	
Emily		45	
Jack		42	
Jane		40	
*Jim		40	
Katherine		40	
LeeAnn		47 or 48	average 47.5
Malcom		35-38	average 36.5
Mic		42-50	average 46
Sheri		40	
*Tamra		45	
Tona		40-50	average 45
Sharon		45	
Jean		47	
Janet		45-48	average 46.5
*Nancy		47.5	

Note. * = sophomore

The average rating the students gave themselves was 40. My initial response was surprise, because this class did not respond the way I thought a non-select choir would. I expected the fact that these students were not in the school's select group would color their sense of competence. However, since the majority of students in this class were freshmen who had yet to try out for a select ensemble, they did not have a negative experience to tarnish their evaluation of their singing.

Some of the students did however experience problems with the cooperative goal structure because they needed more personal feedback about their singing. For example, I observed Bonita requesting feedback from Mrs. T 16 times during the course of the observation period. Other students like Jean and Jane quoted above resented being lumped into the group and having to listen to the teacher discipline others when they themselves were not off task. Also Janet and Kim began to feel personally responsible when the example they tried to set for their classmates did not alter the class dynamic and thus felt bad when a class session did not go as it should have.

Exceptions to the cooperative goal and group reward structure. The exceptions to Mrs. T's predominantly cooperative goal and group reward structure were: a) when she told the students to sing their very best on their first semester performance final because it was also their audition for the select group, b) when she announced and praised the efforts of the students who made the district choir, c) when (in May) she held auditions in class for the two solos written in the music the students were performing for the spring pops concert, and d) when she failed to discipline Emily with the same frequency as she did other students who talked and when she asked Emily (who sat directly in front of her) to run an errand for her (twice). All four of these deviations from the cooperative structure she typically maintained, were competitive in nature.

Student goal citations not expected in a cooperative goal structure. One would not expect to see much ego goal orientation in the students in this class because there was no emphasis placed on students doing better than their classmates (competitive goal structure). The ego goal citations were indeed extremely low and were found mainly among students active in other competitive activities. For example, the following excerpt shows Kim and Janet's reaction to Mrs. T's treatment of Emily:

Mrs. B: Who would you most like to impress of all the people you know?

Kim: Probably the directors at church and Emily. Because she tries out for everything and makes everything.

Mrs. B: OK, your directors at church and Emily. How about you, Janet?

Janet: Mrs. T. Because she has favorites and like if you can impress her you can get something really good.

Mrs. B: When you say she has favorites…

Janet: Like most teachers have favorites, kids they treat good.

Mrs. B: What does she do that you think she has favorites, who are her favorites?

Janet and Kim: Emily

Janet: And Jack… 'cause they're good.

Mrs. B: How do you know they are her favorites what does she do to indicate this?

Janet: I don't know, I guess it's 'cause Emily gets to do a lot of stuff.

Mrs. B: Mrs. T picks her to do stuff?

Janet: Yeah.

Mrs. B: Like what?

Janet: Like solos and stuff.

Mrs. B: Kim you nodded when Janet mentioned favorites. What do you think?

Kim: In the other choirs I know from my sister that it is [name withheld] and [name withheld].

Mrs. B: And do they get special treatment or privileges?

Kim: When it comes to the musical they get leads. So if you're not a lead you don't get anything so if you're not one of her favorites, you don't get anything. [19/K/11]

The focus of the excerpt above seems to be Janet and Kim's accusation that Mrs. T plays favorites in her selection of students to sing solos and have leads in the musical, an extracurricular activity. At the time of this interview, no solos appeared in the literature sung in class. Hence, there was no reason in this context for them to say that Emily got to do solos. Emily did, however, get a slightly bigger role than Kim in the school's musical. So, even though they did not specifically mention the in-class behavior that I considered to be a minor indication of favoritism (Mrs. T ignoring Emily's social talking), Kim and Janet still perceived an inequality in Mrs. T's behavior towards Emily.

This situation is a good example of how the total school environment can affect the context within a particular class. A school-wide practice that stresses ability-focused goals, such as grades, or competitions and rewards that foster ego goals can interfere with classroom-level practices that seek to foster task goals (Maehr & Fyans, 1989; Maehr, Midgley, & Urdan, 1992). Kim was immersed in the school's theater program where she frequently competed for roles and Janet was a serious band student whose director maintained a very competitive environment with such practices as monthly chair challenges performed in class. The influence of these competitive practices could very well have encouraged Janet and Kim to have ego goals in chorus.

Because ego goals predispose students to make ability comparisons, it did not surprise me that Janet and Kim were the only two students to comment in any fashion on Emily's ability or the minor deviation from a group reward structure seen in Mrs. T's behavior towards Emily. Further, it is apparent from Emily, Janet, and Kim's interviews that there was some rivalry among them for Mrs. T's favor.

Summary

Many students discussed their concerns about Mrs. T's cooperative goal and group reward structure in their interviews. As Ames (1984) posited for such settings, the students expressed concern that a) they were perceived as similar in ability and deserving of reward even when their performances varied in quality; b) their behavior was judged by their willingness to put forth effort, but not an individualistic concept of effort, rather a cooperative effort serving the group goals; and c) blame was a noticeable feature in that individuals were punished for not following social norms rather than rewarded for upholding them.

A positive aspect of the first feature that Ames described above (the students will be perceived as similar in ability and deserving of reward even when their performances vary in quality) was a classroom with a very high percentage of students who felt good about their ability to sing.

References

Ames, C. A. (1981). Competitive versus cooperative reward structure: The influence of group performance factors on achievement attributions and affect. *American Educational Research Journal, 18*, 273-287.

Ames, C. A. (1984). Competitive, cooperative, and individual goal structures: A cognitive-motivational analysis. In R. Ames & C. Ames (Eds.), *Research on motivation in education: Student motivation* (pp. 177-207). New York: Academic Press.

Ames, C. A. (1987). The enhancement of student motivation. In D. A. Kleiber & M. L. Maher (Eds.), *Advances in motivation and achievement, 5*. Greenwich CT: JAI Press.

Ames, C., & Felker, D. W. (1979). An examination of children's attributions and achievement-related evaluations in competitive, cooperative, and individualistic reward structures. *Journal of Educational Psychology, 71*, 413-420.

Aronson, E., Stephan, C., Sikes, ZJ., Blaney, N., & Snapp, M. (1978). *The jigsaw classroom.* Beverly Hills, CA: Sage.

Baden, B. D., & Maehr, M. L. (1986). Confronting culture with culture: A perspective for designing schools for children of diverse sociocultural backgrounds. In R. Feldman (Ed.), *Social psychology applied to education* (pp. 289-309). New York: Academic Press.

Casey, P. F. (1994). A status study of nonselective concert bands at selected colleges and Universities (Doctoral dissertation, Ohio State University, 1994). *Dissertation Abstracts International, 54,* (08), A 2936.

Covington, M. V. (1984). The self-worth theory of achievement motivation: Findings and implications. *The Elementary School Journal, 85*, 5-20.

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*, 451-462.

Hill, K. T. (1984). Debilitating motivation and testing: A major educational problem-possible solutions and policy application. In R. Ames & C. Ames (Eds.), *Research in motivation in education: Student motivation* (pp. 245-274). New York: Academic Press.

Johnson, D., Maruyama, G., Johnson, R., & Nelson, D. (1981). Effects of co-operative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89,* 47-62.

Maehr, M. L. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. *Review of Educational Research, 46*, 443-362.

Maehr, M. L. (1984). Toward a theory of personal investment. In R. Ames & C. Ames (Eds*.), Research on motivation in education Vol. I* (pp. 115-144). Orlando: Academic Press.

Maehr, M. L. (1992). Transforming school culture to enhance motivation. Paper presented at the Annual Meeting of the American Educational Research Association: San Francisco, CA, April 20-24.

Maehr, M. L., & Braskamp, L. A. (1986). *The motivation factor; A theory of personal investment.* Lexington, MA: Lexington Books.

Maehr, M. L., & Fyans, L. J. (1989). School culture, motivation, and achievement. In M. L Maehr & C. A. Ames (Eds.), *Advances in Motivation and Achievement: Motivation Enhancing Environments, Vol. 6* (pp. 215-247). Greenwich, CT: JAI.

Maehr, M. L.,& Midgley, C. (1991). Enhancing student motivation: A school-wide approach. *Educational Psychologist, 26,* 399-427.

Maehr, M. L., Midgley, C., & Urdan, T. (1992). School leader as motivator. *Educational Administration Quarterly, 18*, 412-431.

Maehr. M.L., & Nicholls, J. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in cross-cultural psychology, Vol. 3*, (pp. 221-267). New York: Academic Press.

Meece, J., Blumenfeld, P.C., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514-523.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*, 800-814.

Nicholls, J. (1984). Achievement motivation: conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

Differential Expectation Effects as Factors in Evaluations and Feedback of Musical Performance

Mary Ellen Cavitt
The University of Texas at San Antonio

In the last 40 years, experimental research has clearly demonstrated that teacher expectations can have self-fulfilling prophecy effects. Although expectations of teachers concerning students in their own classrooms are generally accurate and based on valid information, teacher expectation effects on student achievement that do occur tend to be undesirable, limiting effects of lowered expectations (Brophy, 1983). Brophy and Good (1970b) found that even though high expectation students succeeded much more often and failed less often, they were more likely than low expectation students to be praised when they did succeed and less likely to be criticized when they failed. Teachers failed to provide feedback to high expectation students only about 3% of the time, but failed to give feedback to low expectation students almost 15% of the time (Brophy & Good, 1970b). Good and Thompson (1998) reviewed research on the communication of performance expectations and found that some teachers "gratuitously praise low-achieving students in ways that indicate low performance expectations; whereas, other teachers communicate low expectations by criticizing low-achieving students disproportionately more often for incorrect answers than is the case for high-achieving students". Other methods of communicating low performance expectations include protecting the student from failure or embarrassment by not giving contingent feedback, criticizing a given student proportionately more often than other students following a wrong answer, praising the student proportionately less often than other students following a correct answer, asking the student to answer only simple questions, and calling on the student only when he or she raises a hand (Good & Thompson, 1998).

A number of studies in music have focused on the relationship between expectations and evaluations of music performances. Duerkson (1972) found that subjects rated performances labeled as "student" performances lower than "professional" even though the recordings were identical. Similarly, Cavitt (1997) found that subjects listening to identical recordings had higher

ratings and expectations for bands labeled "high school" than for bands labeled "beginner". Cassidy and Sims (1991) investigated the effects of special education labels on peers' and adults' evaluations of a youth choir. They concluded that ratings of performances might have been higher because the performers exceeded initially low rater expectations. Schultz (1994) studied the influence of talent expectations and nature/nurture beliefs on evaluations of music performances and concluded that a student who was depicted as talented was rated significantly higher than a student depicted as hard working. Elliot (1995/1996) found that both race and gender influenced judgments of music performances even when those judgments were made by experienced music educators.

The purpose of this study was to investigate the expectation effects of students labeled as high or low ability and/or high or low effort on pre-service music teachers' evaluation of a musical performance.

Method

A compact disc recording of a professional trumpet player performing the first 25 measures of a grade one and a half concert band piece entitled "Fanfare and Fugue" by Anne McGinty was recorded two times on a Marantz Professional Model CDR631 Compact Disc Recorder. The trumpet player was asked to perform with (1) accuracy and good tone on the first performance trial, (2) accuracy, poor tone and without any dynamic contrast on the second performance trial. Each performance was then dubbed twice on a stimulus compact disc for a total of four performances. Pre-service music teachers were told that there were four different seventh grade performers. These performances were reordered and paired with a written anecdote describing the performers' ability and effort. The performances were described in the following order and pairing: Student 1 - recording one was paired with a description of a high ability, high effort student; Student 2 - recording two was described as a low ability, high effort student; Student 3 - recording one was described as high ability, low effort, and Student 4 - recording two was described as low ability, low effort. The first group of the participants (*n*=18) were asked to read the description of the student, listen to the performance, and evaluate the overall performance by circling from among the terms “very poor”, "poor", "fair", "good", and "excellent”. The order of student descriptions of ability and effort were then altered and the other group of participants (*n*=21) was asked to read the student descriptions, listen to the recording, and evaluate performances (See Table 1).

Table 1
Labels and Performances Used as Stimuli (In Order of Presentation)

Group A *n*=18		
Order number	Label	Performance Heard
1	High Ability/High Effort	Good

2	Low Ability/High Effort	Poor
3	High Ability/Low Effort	Good
4	Low Ability/Low Effort	Poor

Group B *n*=21

Order number	Label	Performance Heard
1	High Ability/Low Effort	Good
2	Low Ability/Low Effort	Poor
3	High Ability/High Effort	Good
4	Low Ability/High Effort	Poor

Results

Participant's evaluations of overall performance were coded as 1= "very poor", 2= "poor", 3= "fair, 4= "good", and 5= "excellent". The data set was analyzed using a repeated measures analysis of variance (ANOVA) with one between-subjects factor (the *order* that performances were labeled) and two within-subjects factors (*good or poor* performances and performances labeled *high effort or low effort*). As would be anticipated, there was a significant difference between how participants rated the good and poor performances [F (1,37) = 25.77, p< .001]. More importantly, there was a highly significant difference in the way participants rated performances when they were labeled as a high effort or low effort student [F (1,37) = 148.50, p<. 001]. This is of great interest because, as you recall, participants may have been listening to the same performance with the only difference being that of the label. There was also a significant interaction between the good/poor performances and the high/low effort labels [F (1,37) = 4.77, p <. 035]. This indicates that although these factors are not independent of each other, the label had a significant effect on how participants rated students. The labels did not yield an effect independent of the good or poor performances. The order in which students were labeled had no significant effect on how participants rated performances. Table 2 shows the mean ratings for performances irrespective of the order presented.

Table 2
Means for Evaluation Ratings of Labeled Student Performances

Performance	Label	Mean

Good	High Effort	4.08
Good	Low Effort	3.87
Poor	High Effort	3.05
Poor	Low Effort	2.64

This table indicates clearly that when pre-service teachers heard the same performance and the student was labeled as "low effort" they evaluated the student lower.

Discussion

The purpose of this study was to investigate the expectation effects of students labeled as high or low ability and/or high or low effort on pre-service music teachers' evaluation of a musical performance. Participants' responses indicated that labels did have a significant effect on the way they evaluated the student.

Results indicated that the interaction between performances labeled as high/low effort didn't qualify either of the main effects. There was a smaller difference between the means of the students labeled as "Good/High Effort" and "Good/Low Effort" versus the difference in means between the "Poor/High Effort" and the "Poor/Low Effort" players. This may indicate that effort is perceived as being less important if a student plays well than if a student plays poorly.

Teachers' beliefs about students' abilities are important. Previous studies in music have demonstrated that the expectation effects we have for students can have an effect on the way we evaluate and teach students. As discussed previously, some of the labels teachers place on students may be based on valid, accurate information while others may be based on unreliable biases. Differential expectation and treatment may be detrimental to student performance. Emphasizing student potential rather than deficits and systematically developing positive and higher expectations for students may lead to more effective classrooms and more successful students.

References

Asmus, E. (1986). Student beliefs about the causes of success and failure in music: A study of achievement motivation. *Journal of Research in Music Education, 34* (4), 262-278.

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75* (5), 631-661.

Brophy, J. & Good, T. (1970b). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61* (5), 365-374.

Brophy, J. & Good, T. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart & Winston.

Cassidy, J. & Sims, W. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education, 39* (1), 23-34.

Cavitt, M. E. (1997). Effects of expectations on evaluators' judgments of music performance. *Texas Music Education Research, 1997*.

Cooper, H. & Tom, D. (1984). Teacher expectation research: A review with implications for classroom instruction. *The Elementary School Journal, 85* (1), 77-89.

Duerkson, G. (1972). Some effects of expectation on evaluation of recorded musical performance. *Journal of Research in Music Education, 20* (2), 268-272.

Elliot, C. (1995/1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education.*

Good, T. & Thompson, E. (1998). Research on the communication of performance expectations: A review of recent perspectives. In J. Brophy (Ed.), *Advances in research on teaching: Vol. 7. Expectations in the classroom.* Greenwich, Conn: JAI Press.

Haskett, M. (1968). An investigation of the relationship between teacher expectancy and pupil achievement in the special education class. *Dissertation Abstracts, 29,* 4348A-4349A.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93* (4), 429-445.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57* (3), 469-480.

Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review, 8*, 193-210.

Raudenbusch, S. (1984). Magnitude of teacher expectancy effects of pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76* (1), 85-97.

Rist, R. (1970). Student social class and teacher expectations: the self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 40 (3)*, 411-451.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (2nd ed.). New York: Irvington.

Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* New York: Holt, Rinehart and Winston.

Schultz, J. (1994). *The influence of talent expectations and nature/nurture beliefs on evaluation of music student performances*. Ann Arbor, MI: University Microfilms International.

West, C. & Anderson, T. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research, 46* (4), 613-630.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

Perception of Rubato by Non-Music Majors, Music Majors, and Piano Majors of Chopin's *Fantasie Impromptu, op. 66*

Allison Hudak
The University of Texas at Austin

Detecting rubato is a task that highly skilled musicians do in any musical situation. Detecting rubato has been discussed as something that may only be perceived by those with finely acute hearing. Some argue that less experienced musicians cannot detect rubato. There has been research that suggests that rubato can be taught to music majors in order to create more musical performances (Johnson, 1998). Johnson investigated how instruction in the use of specific rhythmic nuances influenced the performed timings of a musical performance. Volunteer participants were asked to learn and musically perform an excerpt from Mozart's *Concerto for Horn and Orchestra, No. 2*. The subjects used the computer software entitled "Instant Pleasure" to record their responses. No prior musical training was required to perform music using the program. The only element that the operator controlled was the rhythmic onset of each pitch. Participants were then taught the rhythmic tendencies that professional musicians used in the excerpt. Participants created a second recording using the newly taught information. Results indicated that subjects did use significantly more rubato in their posttest performance and that usage more closely reflected the model performance (Johnson).

Other studies have found that the music educational level does impact listeners' perception of rubato in music. In a study conducted by Sheldon and Gregory (1997), similarities and differences in how listeners with different levels of educational experience demonstrate perception of tempo modulation in music using the Continuous Response Digital Interface (CRDI). Subjects' responses were significantly different among degrees of tempo change in increasing and decreasing examples. Significant differences were also found among subject groups for both increasing and decreasing examples. Significant interactions were found between education level and tempo change for increasing and decreasing examples (Sheldon & Gregory, 1997).

Researchers have concentrated on the issues of the perception of tempo and beat-rate change. Variables associated with tempo modulation study cover a broad range. Some researchers have extracted beat speed out of a music context using monotonic stimuli (Brittin, 1993; Drake 1968; Duke, 1989a; Duke, Geringer & Madsen, 1991). Some investigators maintain that study of beat modulation within a musical context is viable (Duke, 1989a; Duke, Geringer, & Madsen, 1988).

Other investigations have focused on detecting tempo change when the melody stands alone and how it may affect listeners' ability to detect rubato (Wang, 1984). Research indicates that rubato changes are more readily detected when there are more notes per beat in the musical excerpt.

Madsen, Duke, and Geringer (1984) studied pitch as a variable in tempo change detection and determined that concomitant pitch and tempo variation in band music excerpts may have an effect on listeners' capability to detect changes among those factors. In another study, Duke, Geringer, and Madsen (1988) found that tempo alteration seemed to have a greater effect on pitch response than did pitch alteration on tempo.

Although most tempo discrimination studies use audible stimuli, a few investigations have involved the use of visual and kinesthetic stimuli as well. Brittin (1992) discovered that musicians are more accurate in detecting tempo deceleration compared to acceleration when the stimulus is a visual, conducted beat. The opposite results were found with non-musicians. In another study, Sheldon (1994) found that subjects were more accurate when listening only or listening and physically maintaining a beat compared to listening and watching a conductor.

In regard to musical training, Miller and Eargle (1990) argued that there was no strong evidence from other research that musical training by itself played a major role in tempo discrimination. Failures to find a relationship between musical training and tempo discrimination had been reported by Geringer & Madsen (1984) and on two occasions by Wang (1983, 1984). Madsen (1979) using a metronome, did find a difference favoring musicians during part of his experiment. In an experiment by Miller and Eargle (1990), children ages 7 to 11 and adults were asked to detect changes in an unaccompanied drumbeat that increased, decreased, or maintained the same tempo across measures. Marked differences in performance as a function of age were found for the change trials, older subjects being more accurate, while musical training was relatively unimportant. For trials in which the beat remained constant, however, subjects with musical training were more accurate, while age, by itself, had little effect upon performance.

In current studies, researchers have found different results regarding musical training. They have focused on tempo detection among varying ages and levels of musical training (Brittin, 1992, 1993; Duke, 1989b; Duke, Geringer, & Madsen, 1991). Generally, subjects with more training are better at tempo-change assessment than those with less training are. Musicians often respond to tempo change more quickly than do non-musicians. However, in many studies, data regarding accuracy in rubato detection with respect to the direction of the tempo modulation is contradictory. Some research suggests music majors and non-majors detect tempo decrease more accurately than the detect tempo increase (Kuhn, 1974; Madsen, 1979), whereas other research indicates the opposite (Geringer & Madsen, 1984; Madsen, Duke, & Geringer, 1984; Wang, 1983; Yarbrough, 1987). Possible reasons for these contradictory findings could be related to the stimuli used in the experiments. Such stimuli include: metronomic clicks versus examples heard in a musical context, length of the example, subdivisions as determined by the rhythmic content of the example, subjects' predetermined tempo preference for certain musical pieces, and listeners' general preference for fast tempos.

In a study conducted by Johnson (1996a), results indicated significant differences for musicians compared to non-musicians regarding rubato assessments. Musicians agreed with expert assessments, whereas non-musicians' scores seemed haphazard. The purpose was to investigate musicians' and non-musicians' assessment of perceived rubato in musical performance. The study suggested that rubato was an extremely subtle musical nuance. Though non-musicians and less proficient musicians seemed to be able to assess performances on somewhat more obvious variables, it seemed that only the most proficient musicians could evaluate the usage of rubato in performance. Based on the data from this study, it appeared that there was a relationship between musicianship and the use of rubato and that use of rubato might be one element that separates the very finest performance from the ordinary.

Wang and Salzberg (1984) administered a tempo discrimination task to string students ages 7 to 18. Tempo in musical excerpts increased, decreased, or remained constant, and listeners were asked to note both the point and direction of tempo change. A curvilinear relationship was found between accuracy and age as well as between accuracy and years of musical training. Performance improved until age 13 after which it decreased.

Method

This study was designed to determine if listeners from varying levels of education perceive rubato differently. Subjects listened to a recording of the first seventy measures of *Fantasie Impromptu*, op. 66 (performed by Vladimir Ashkenazy) and marked an "x" on every measure where perceived rubato occurred on their printed scores. Further analysis was conducted to determine if subjects more readily detected rubato in the *allegro agitato* section or the *largo* section in the piece.

Four groups of ten subjects were chosen for this study (N=40). Subjects were selected from the University of Texas at Austin class piano sections: two groups of non-music majors, one group of music majors (non-pianists), and one group of piano majors (ranging from undergraduate to graduate students). All groups of students were given the same definition of rubato, instruction on how to mark the score, and told how to follow the measures. Below is documentation of what was read to each group of students.

> **Rubato** (Harvard Dictionary, 1986): Rubato is an Italian term meaning stolen time. In performance, rubato is the practice of altering the relationship among written note-values and making the established pulse flexible by accelerating and slowing down the tempo; such flexibility has long been an expressive device.
> **Rubato** (Author's Definition): In other words, if you speed up or slow down, you are using some form of rubato. If you can snap your fingers to the steady beat, then you are not using rubato.
> **Instruction**: You will hear seventy measures of piano music as indicated on your score. If you detect any rubato that the performer uses on the recording, mark an "X" through that measure. I will count each measure number beginning on the downbeat. You will notice that the measure numbers are written at the beginning of each line. Use those as a guide.

All subjects were asked if they had any questions about rubato. Then all students were given the musical score of the first seventy measures of *Fantasie Impromptu, op. 66*. Students were provided pencils to mark "X's" through measures. All subjects heard the recording once. After subjects marked their scores, the author collected the scores.

In order to determine the measures that exhibited rubato, a panel of three doctoral piano students listened to the recording three times and marked individual scores of their perceived rubato. After the listening portion, panel members discussed which measures were chosen. The percentage of agreement was: Section A (*allegro agitato)* = 96.46% and Section B (*largo)* = 92.85%. This was calculated for both the A and B sections of *Fantasie Impromptu, op. 66*, measures 1-70.

Measures were not counted as "correct" if a subject marked an "incorrect" rubato or if the student failed to mark a rubato where one occurred. After all the measures were marked as "correct" or "incorrect," the author counted all the correct measures and used those scores to calculate the similarities and differences between the groups.

Results

All data scores were analyzed using a Two-Way Analysis of Variance with one repeated measure and a post hoc Duncan test. Results indicated that there was a significant difference between pianists' scores versus the other three groups in the study. Pianists' scored more closely to the panel of experts than any other group. The following table shows the mean ranks of rubato detection:

Table 1.
Mean Ranks of Perceived Rubato in Chopin's Fantasie Impromptu (mm. 1-70).

Non-Music Majors Group 1	Non-Music Majors Group 2	Music Majors (non-pianists)	Piano Majors
M	*M*	*M*	*M*
23.45	23.70	25.35	*28.10

Note. Results indicated that there was a significant difference between the piano majors' scores in their perception of rubato versus the other three groups were $p<0.01$.

Although the music majors seemed to score higher than the other two groups of non-music majors, there was not a significant difference between their scores and the two groups of non-music majors. These results do seem to suggest that the amount of piano education does effect how rubato is detected when analyzing a piano score and listening to a piano recording.

Another Two-Way Analysis of Variance was conducted on the data in order to determine if there was a difference in detecting rubato more easily in either the A or B sections of the music. Results indicated that there was a significant difference in detecting rubato more readily in the A section rather than the B section.

Table 2.
Mean Ranks of the Allegro Agitato and Largo Sections in Fantasie Impromptu for All Subjects Combined.

Allegro agitato (mm. 1-40)	Largo (mm. 41-70)
M	*M*
*29.53	20.76

Note. Results indicated that there was a significant difference between detecting rubato in the *allegro agitato* section more readily than the *largo* section where $p<0.01$.

Results indicated that there was no significant difference between the interaction of groups in detecting rubato in section A or section B. Results indicated that there was more agreement of where rubato occurred in the *allegro agitato* section rather than the slower *largo* section. This supports other research indicating that it is easier to detect rubato in faster passages of music rather than slower sections of music (Geringer & Madsen, 1984; Madsen, Duke, & Geringer, 1984; Wang, 1983; Yarbrough, 1987).

Discussion

This study was designed to examine if listeners from varying levels of music education perceived rubato differently. A secondary purpose was to examine whether listeners could detect rubato more accurately in either fast or slow passages of music. Pianists scored significantly higher in detecting rubato in the musical score in comparison with non-music majors and music majors (non-pianists). Subjects scored differently in sections A and B of the piece, having more agreement in the A (*allegro agitato*) section. Subjects had less agreement in detecting rubato in the B (*largo*) section of the music.

All subjects involved in this study were involved in some form of group piano class. All subjects have experienced some form of rubato in either their own solo repertoire or in class discussions. Both sections of non-music majors had a special lesson on rubato a few weeks before this study was conducted. Subjects were encouraged to listen for rubato when they attended mandatory concerts. The music majors may not be proficient pianists, but on a day-to-day basis, they are confronted with rubato in their solo literature or in ensembles. It is interesting that the music majors (non-pianists) did not score significantly higher in comparison with the non-music majors. This could be due to several factors. First, it is possible that the music majors heard subtle rubato in measures that the expert panel did not mark. If subjects in this group marked that they had heard rubato in a particular measure and it did not match the panel sheet, then those measures were counted as "incorrect." This could have caused the music majors' scores to decrease when in reality they heard rubato, but marked either the wrong measure, or marked too many measures of perceived rubato compared to the "experts."

When comparing the non-music majors, they seemed consistent between the two groups. It is obvious that their lack of experience in detecting rubato produced lower scores. This may be an effect of the difficulty in following the score. The non-music majors are not used to looking through a difficult piece of piano literature. For many, this could have been the first time looking at such difficult music. The difficulty of the music may have affected their capability to focus on the music. Perhaps many of them felt overwhelmed by looking at such a difficult piece of music. It is plausible that subjects could not follow the score even though the measure numbers were given to them for each measure. Again, if subjects marked a measure as rubato and it did not match the panels' answers, then the measure was counted as incorrect. It is possible that the subjects heard rubato, but got lost in the score and could not mark the exact measure because of confusion.

It is not surprising that the pianists scored significantly higher than the other groups. Pianists were looking at a piano score. All pianists were proficient on the piano and have had many years of experience performing and interpreting rubato. But it is interesting that there was a significant difference between their scores versus the music majors (non-pianists) scores. As musicians, it seems as though they should all be able to detect rubato in any genre. Perhaps familiarity with an instrument has an influence on listeners' ability to detect rubato.

Further investigation of rubato detection seems warranted. If this study were replicated, it would be beneficial to play the excerpt for all students three times in order for them to check over their responses. Perhaps subjects needed a "warm-up" experience listening for rubato before marking their perceptions in the printed score. Having the opportunity to check over their work may have yielded more accurate responses concerning rubato.

These findings seem to support other studies conducted on rubato and tempo modulations. In Sheldon and Gregory's study (1997), the level of educational experience affected the perception

of tempo modulation as demonstrated with a CRDI. Their results also suggested that students were more likely to correctly hear dragging over rushing.

When reviewing the passages of *Fantasie Impromptu,* subjects in this study also tended to detect measures of tempo deceleration rather than measures using tempo acceleration. Other studies have supported this natural occurrence that listeners respond more quickly and with greater sensitivity to tempo decrease compared to tempo increase (Brittin, 1992; Kuhn, 1974; Madsen, 1979; Wang, 1984), while it is inconsistent with the findings of Sheldon (1994). More experienced listeners tend to hear tempo changes differently (Brittin, 1992, 1993; Duke 1989b; Duke, Geringer, & Madsen, 1991; Miller & Eargle, 1990). Sheldon and Gregory (1997) discussed that although all groups demonstrated differing degrees of tempo perception mainly commensurate with the magnitude of tempo change, perceptions of the more experienced listeners seemed to be somewhat set apart from groups with less experience.

This study focused subjects' attention on the printed score when detecting rubato. Subjects had the opportunity to detect the slightest amount of rubato by marking the printed score. This procedure had not been used in previous studies. It may be beneficial to allow students as much time as they need in order to review where rubato occurs in the score. It may have been difficult to make an accurate assessment on one listening episode. It may be interesting to place the score on a Finale computer program that allows the listener to follow the score with a pointer. Perhaps the program SCRIBE (Duke & Farra, 2001) could also be used. SCRIBE (Simple Computer Recording Interface for Behavioral Evaluation) would allow the researcher to measure the frequency and duration of the rubato. Subjects could follow the score on the computer so they would not "lose their place" while keeping their finger on a mouse enabling them to click when rubato occurs. This may yield more accurate results.

References

Brittin, R.V. (1992). Perceptions in conducting: Accuracy in detecting modulated beat. *Bulletin of the Council for Research in Music Education, 113*, 1-8.

Brittin, R.V. (1993). Discrimination of aural and visual tempo modulation. *Bulletin of the Council for Research in Music Education*, *116*, 23-32.

Chopin, Frederic. *Favourite Piano Works: Beliebte Klavierwerke.* "*Fantasie-Impromptu, op. 66.*" Performed by Vladimir Ashkenazy. The Decca Record Company Limited: London, 1995.

Drake, A.H. (1968). An experimental study of selected variables in the performance of musical durational notation. *Journal of Research in Music Education*, *16*, 329-338.

Duke, R.A. (1989a). Musicians' perception of beat in monotonic stimuli. *Journal of Research in Music Education*, *37*, 61-71.

Duke, R.A. (1989b). Effect of melodic rhythm on elementary students' and college undergraduates' perceptions of relative tempo. *Journal of Research in Music Education, 37*, 246-257.

Duke, R.A., & Farra, V. (2001). SCRIBE: Simple Computer Recording Interface for Behavioral Evaluation [Computer software]. Austin, TX: Learning & Behavior Resources.

Duke, R.A., Geringer, J.M., & Madsen, C.K., (1988). The effects of tempo on pitch perception. *Journal of Research in Music Education, 36*, 108-125.

Duke, R.A., Geringer, J.M., & Madsen, C.K. (1991). Performance of perceived beat in relation to age and music training. *Journal of Research in Music Education, 39*, 35-45.

Duke, R.A., & Pierce, M.A. (1991). Effects of tempo and context on transfer of performance skills. *Journal of Research in Music Education, 39*, 93-100.

Geringer, J.M., Duke, R.A., & Madsen, C.K. (1992). Musicians' perception of beat note: Regions of beat change in modulating tempos. *Bulletin of the Council for Research in Music Education*, *114*, 21-33.

Geringer, J.M., & Madsen, C.K. (1984). Pitch and tempo discrimination in recorded orchestral music among musicians and non-musicians. *Journal of Research in Music Education, 32*, 195-204.

Johnson, C.M. (1996a). Musicians' and nonmusicians' assessment of perceived rubato in musical performance. *Journal of Research in Music Education, 44*, 84-96.

Johnson, C.M. (Fall 1998). Effect of instruction in appropriate rubato usage on the onset timings and perceived musicianship of musical performances. *Journal of Research in Music Education, 46* (3), 436-445.

Kuhn, T.L. (1974). Discrimination of modulated beat tempo by professional musicians. *Journal of Research in Music Education, 22*, 270-277.

Kuhn, T.L. (1987). The effect of tempo, meter, and melodic complexity on the perception of tempo. In C.K. Madsen & C.A. Prickett (Eds.), *Applications of research in music behavior* (pp. 165-174). Tuscaloosa: University of Alabama Press.

Madsen, C.K. (1979). Modulation beat discrimination among musicians and nonmusicians. *Journal of Research in Music Education, 27*, 57-67.

Madsen, C.K., Duke, R.A., & Geringer, J.M. (1984). Pitch and tempo discriminations in recorded band music among wind and percussion musicians. *Journal of Research in Music Education, 20*, 20-29.

Miller, L.K., & Eargle, A. (1990). The contributions of development versus musical training to simple tempo discrimination. *Journal of Research in Music Education*, *38*, 294-301.

"Rubato." *The New Harvard Dictionary of Music.* Edited by Don Randel. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1986.

Sheldon, D.A. (1994). Effects of tempo direction, musical experience, and differential modes of listening on the perception of modulating tempo. *Journal of Research in Music Education*, *42*, 190-202.

Sheldon, D.A. & Gregory, D. (1997). Perception of tempo modulation by listeners of different levels of educational experience. *Journal of Research in Music Education*, *45*(3), 367-379.

Wang, C.C. (1983). Discrimination of modulated music tempo by music students. *Journal of Research in Music Education, 31*, 49-55.

Wang, C.C. (1984). Effects of some aspects of rhythm on tempo perception. *Journal of Research in Music* , *32*, 169-176.

Wang, C.C. & Salzberg, R.S. (1984). Discrimination of modulated music tempo by string students. *Journal of Research in Music Education*, *32*, 123-131.

Yarbrough, C. (1987). The effect of musical excerpts on tempo discriminations and preferences of musicians and non-musicians. In C.K. Madsen & C.A. Prickett (Eds.), *applications of research in music behavior* (pp. 175-189). Tuscaloosa: University of Alabama.

Edited by
Mary Ellen Cavitt, The University of Texas at San Antonio

An Analysis of Certification Practices for Music Educators in the Fifty States

Michele Henry
Baylor University

Because each state determines its own standards for certifying teachers, practices among states vary significantly. Differences in certification practices appear even more varied when looking at a single certification area, such as music. In 1972, Wolfe compiled a detailed account of state certification practices for music educators. Erbes (1984) replicated this study, noting the changes that had occurred during the preceding decade. Among the changes in certification that had transpired was a decrease in the number of states offering K-12 certification in music. While many states continued to grant an all-encompassing license for teaching music at any grade level, other states had begun to issue individual certifications for various grade-level groupings. Additionally, thirteen states implemented some form of required testing for certification in 1983, an increase of ten states from the 1972 study. In 1984, eighteen states required testing. In a subsequent report, Erbes (1987) identified 29 states that required testing to receive initial certification, with 13 of the states requiring testing of music content knowledge.

In comparison to Wolfe's 1972 findings, Erbes (1984) also reported a decline in the number of states offering life certification, which coincided with the development of state-mandated continuing education programs and advanced certification requirements. Rowls and Hanes (1982) identified 27 states requiring recertification for teachers. Nine states reported no recertification requirements, either renewing certificates automatically after a requisite number of teaching years or issuing lifetime certification. A subsequent report on the teaching profession (Holmes Group, 1986) recommended the establishment of a three-tiered certification system for educators, initially issuing a non-renewable certificate to all entry-level teachers who successfully passed rigorous basic skills and content area exams. The Carnegie Task Force on Teaching as a Profession (1986) called for the development of a national certification for teachers at two levels—standard certification but at a more stringent level than previously required, and advanced certification attainable only by exceptional teachers. Erbes (1987) predicted that the ideas put forth by these two reports would both be in place by 2000.

An awareness of various state practices for granting certification can be valuable information for a variety of persons. Bruner (1977) observed "Americans are a changing people; their geographical mobility makes imperative some degree of uniformity among high school and primary schools. Yet the diversity of American communities and of American life in general makes equally imperative some degree of variety" (p. 9). These observations also pertain to teacher credentialing practices in the United States. A mobile society will see not only students

relocating from state to state, but teachers as well. While uniform teacher certification procedures, as called for by Gallegos (1978), may seem accommodating to relocating teachers, the diversity of school settings and student populations among the states necessitates variety in certification practices as well.

In as much as experienced teachers may choose to relocate to another state during some portion of their teaching careers, information about reciprocity agreements between the states in question can have a significant effect on relocation plans. While some states honor reciprocity agreements such as the National Association of State Directors of Teacher Education (NASDTEC) Interstate Contract in full, others may not recognize licensure from any other state. Still others grant temporary certification while deficiencies, such as required tests or state mandated courses, are eliminated.

College students and prospective college students benefit from knowing state certification requirements. Because education degrees—in particular music education degrees—have continued to increase in size, advanced planning on the part of the college student and his/her family is essential. Often, music education students are not able to complete degree requirements during four years, necessitating either summer school or an extra semester or year of coursework. An awareness of state certification requirements could impact how students choose to spend this valuable time. In states with multiple music certificates, planning to take an additional course or two may permit an additional endorsement on the teaching certificate. In a 1984 article articulating the advantages and disadvantages of double majoring, Duling stated, "Though certification requirements vary from state to state, I believe it would be helpful for music students to study the requirements in their state during their college career to ascertain how they might at least begin selecting their courses in a second area of interest in education" (p. 54). This recommendation applies equally as well to students going to school out of state, particularly if the student plans to return to his/her home state to teach after graduation.

The process of becoming certified to teach is without exception a costly proposition, in terms of investment of finances, time, and effort. Many states require substantial fees for application, fingerprinting, and renewal of certificates. States also require several weeks, or even months, to process certification requests. Additionally, most states also require successful passage of at least one—and as many as four—standardized test(s). These tests are typically offered only at specified times of the year, and results of these tests can also take several weeks to report. This amount of investment merits both careful study and planning on the part of the prospective teacher.

Further, teachers and administrators in higher education also benefit from knowledge of the varying practices for certification around the nation. Very often, these individuals serve as advisors for students who intend to relocate to another state after graduation. Advice on the certification practices of these states, or even methods to obtain this information, could be extremely valuable to these students. Students may choose not to take the required certification tests in the college's state, but rather opt to take only those tests required by the state where they intend to teach. Additionally, awareness of out-of-state certification and teaching standards can shape assignments and activities within methods courses. While students planning to reside and teach in the same state as the college or university may use state goals and standards when planning lessons and identifying lesson objectives, students planning to teach in other states may benefit from citing standards or criteria from other states—or even the National Standards for Music Education (CNAEA, 1994).

Finally, officials responsible for determining certification standards ought to be interested in the practices of their peer states. Although policy-making should not rely upon unproven trends,

the rationale behind changes made in other states' certification procedures may challenge existing assumptions in states considering changes themselves. No certification system is perfect, and every program requires ongoing assessment. The very nature of our educational systems requires that certification practices try to maintain currency with changing educational theory. It then behooves certification policymakers to evaluate the benefits of other models of granting certification. While revisiting and adapting certification policy may manifest itself through reconfiguring certification practices within the existing structure, it may also result in substantially different paths to certification.

In his 1987 article, Erbes cited the development of alternate routes to teacher certification, primarily due to declining numbers in the profession. Since that time, many states have adopted alternative certification programs for individuals who hold college degrees but are not graduates of an approved education curriculum. Berry (2001) identified 41 states offering alternate certification, fourteen of which have been developed in the last two years. Although these programs have many advocates (Finn & Madigan, 2001), there are also many vocal opponents to the idea of so-called "shortcuts" to teacher certification (Berry, 1994; Etheridge, 2000-2001). Regardless of one's opinion on the relative merits of such programs, their increasing commonality has tremendous implications for music education and teacher training at large.

Regardless of the reasons, access to state teacher certification information is an important and necessary fact for hundreds of thousands of educators each year. States such as Connecticut, Kentucky, Minnesota, and Washington each report receiving up to 28,000 inquiries about certification annually: (www.state.ct.us/sde, www.kde.state.ky.us/otec/cert, www.educ.state.mn.us/licen, www.k12.wa.us/cert).

Currently, every state maintains a Web site dedicated to teacher certification information. Formats and accessibility of these Web sites vary considerably. Although some sites are better than others, none is able to provide comprehensive information about certification. Requesting information by phone can be both time-consuming and expensive. Receiving materials by mail can take even longer, and once received may still require explanations by phone, fax, or email. While increasing access to information concerning teacher certification is available, a comparison of practices among states still remains very difficult and time consuming.

Method

The purpose of this study was to compile relevant information for music educators about the current certification practices of each state in the United States, and to conduct an analysis of the commonalities and differences among the states' policies. The results of such an analysis may reveal trends in certification procedures across the nation, or among states with certain similar characteristics. It is not the intention of this study to recommend particular certification structures or requirements.

Although states' terms for their teaching credentials vary—the use of certificate, license, and credential are not interchangeable in many states—for the purposes of clarity in this study, the terms certificate and certification are used to designate the legally-sanctioned document permitting employment in education, regardless of the term designated by each state.

Research questions

This study sought to provide information concerning certification of music educators in each of the fifty states and the District of Columbia. In order to gather comparable information, specific questions guided the information-gathering process for each state's music certification practices. Several of these questions were derived from Erbes report (1984). Additional

questions related specifically to the accessibility of information, for which some avenues were not available in the 1980's. In order to determine the accessibility to certification information, the following questions were asked: (a) What is the name of the governing body for teaching certification in the state? (b) What is the agency phone number that connects to a live person? (c) What is the Web site address for teacher certification in the state?

In order to determine the credentialing structures and processes of each state, answers to the following questions were sought: (a) What content areas are included under certification in music teaching? (b) What are the age-level designations for certification in music teaching? (c) How recent are the current certification practices in the state? (d) What tests, if any, are required for certification in music teaching? (e) Does the state have reciprocity for certification with any other states? (f) What are the types of certificates available and length of validity for the various certificates? (g) Is there an alternative certification program available for those without education degrees? (h) What fees, if any, are required for certification in the state? (i) Are application forms and instructions available on line?

A further goal of this study was to provide an overall analysis of the certification practices identified through the data gathering process. After gathering the above information for each jurisdiction, the following questions were considered: (a) Is there a trend among the various states regarding age-level designations and/or content areas? (b) Is there a trend among the various states regarding levels of certification and/or length of certificate validity? (c) Is there a trend among the various states in requiring certain kinds of tests for certification? (d) What is the level of cooperation among the states in acknowledging teaching certification from other states (reciprocity)? (e) Is there a trend among the various states regarding alternative certification? (f) What is the range of certification fees among the various states? (g) How accessible is information concerning certification in each of the states?

Data collection procedures

Weible and Dumas (1982) conducted a survey of course requirements for certification in the 50 states. Forty-five states provided written responses. Information for the remaining five states was collected by phone. Additionally, any confusing or contradictory information from written responses was clarified by phone.

Data for the current analysis was gathered using a variety of means. Initially, the investigator used an on-line search engine to find a listing of state departments of education (or comparable agencies). Using the links identified by the search engine, the researcher was able to access the appropriate state agencies for all 50 states. Data concerning each question was gathered using only information provided on line.

After obtaining all information available on line, the researcher contacted each of the agencies by phone. All data found online was verified by an official representative of the state credentialing agency. If a discrepancy existed between the information found on line and that provided on the phone, it was noted in the results section of this document. The researcher also requested any information needed for the study not found on the state's certification web site at this time. Finally, the researcher requested hard copies of certification materials and applications from each state. The hard copies of the certification materials were used in effort to elucidate the correct information concerning the afore-mentioned discrepancies or to provide any information that was not gained through the Web searches or phone interviews.

Results

This section addresses research questions asked regarding trends in certification practices among the states. Each question is addressed individually using data gathered from each of the states. Similarities as well as contradictions in philosophies concerning certification are identified.

Age-level and subject area

Forty-four states offer all-level certification for music teachers. The lowest age-range included in all level certification varies from preschool (or nursery), kindergarten, or first grade among states. Thirty states offer only all-level certification in music. Of the 20 states that provide restricted age-level certification, all but four states offer music certification separately at the elementary level. The grade levels included in the elementary range vary, but include combinations from preschool through ninth grade, with K-6 and K-8 being the most common. Nineteen of the 20 age-restricted certification states offer music teaching at the secondary-level as an option. Only Alabama specifies an elementary-only music teaching credential without also offering a secondary-only music credential. Only two states have age designations specifically geared toward middle school, although many states have overlapping certification ranges that include middle school with either elementary certification or secondary certification (see figure 1).

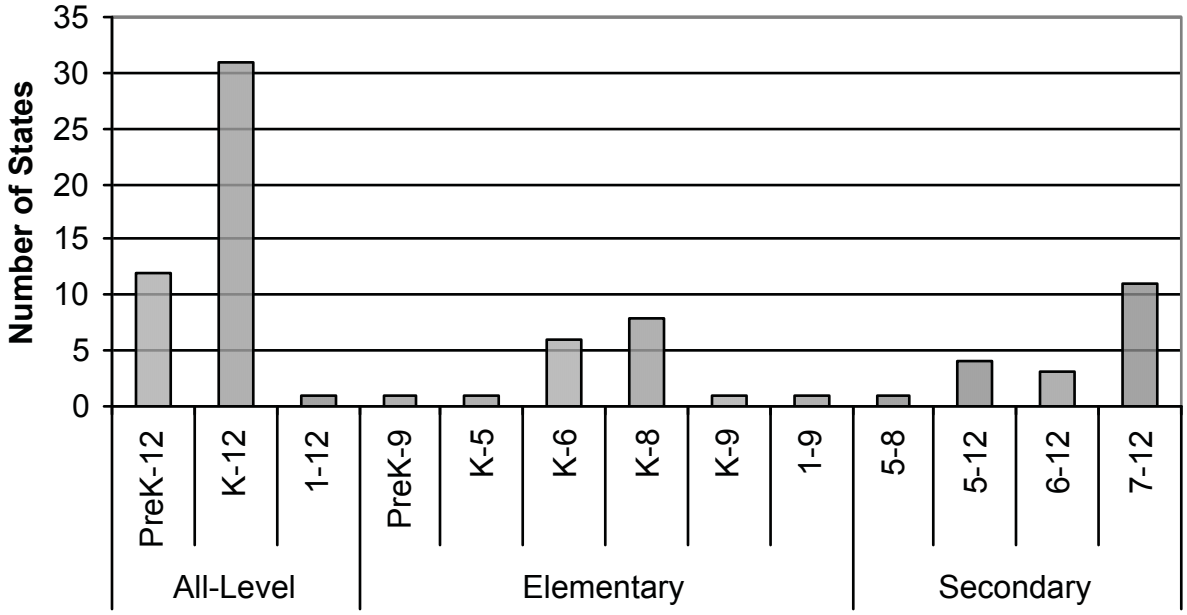

Figure 1. Certification age levels for music in the fifty states.

Thirty-one states consider music a single subject area for certification purposes. Five additional states offer a composite certification for all music areas. Fifteen states differentiate between vocal and instrumental music for certification purposes. In most of these states, either

one is able also to teach general or classroom music. Three states separate certification between vocal, instrumental, and general music. Finally, South Carolina has certification areas for choral, instrumental, piano, violin, and voice.

The age ranges and subject areas for music certification vary greatly among the states. Nineteen states offer only one certification for music, encompassing all grades and disciplines within music. States such as Kentucky explain their rationale behind broader certification as an attempt "to reduce and streamline the credential system to allow greater flexibility in staffing local schools while maintaining standards for teach competence" (KRS 161.028[g]). These broader certifications also imply more responsibility for schools at the local level. "As the state becomes less prescriptive as to who can teach which courses, it becomes increasing important for school districts and school councils to carefully review the transcripts and experiences . . . of prospective teachers. Ensuring a good 'match' between each faculty member's responsibilities and expertise is imperative . . ." (www.kde.state.ky.us/otec/cert). In contrast to states offering broad credentials for music teaching, other states have chosen to segment their credentials to reflect specific populations or disciplines. The Indiana Professional Standards Board voted to establish developmental level standards that reflect the school setting in which teachers are likely to be employed. "The developmental levels, for licensing purposes, need a P-12 connection and should respect school configurations at the local level, while ensuring that educators will be thoroughly prepared for the developmental level which they will teach." (www.in.gov/psb/licensing). Further, states such as Indiana separate music teaching licensure into vocal/general and instrumental general. "The standards clearly define each of the fine arts . . . as a discrete discipline." (www.in.gov/psb/licensing). Although manifesting itself in very different forms, the motivation for these states when constructing their credentialing categories seems to place an emphasis on local control and appropriate matches for educators and classrooms.

Levels of certification and certificate length

In 1984, Erbes identified eight states with entry-level programs established or in development. The Holmes Group (1986) and the Carnegie Task Force (1986) each called for tiered certification systems for educators. Currently, thirty-four states require some type of provisional certification for entry-level teachers. Some states also use this initial certification document for teachers coming from other states with deficiencies to address before full certification is granted. Sixteen states do not distinguish between levels of certification, using only a single credentialing designation. The validity length of initial certificates varies from two years to six years. The validity length for standard certificates varies from three years to lifetime certification. Currently, only five states grant lifetime certification.

The perceptible trend in certificate structure seems to be toward encouraging continuing education for teachers by requiring certificate renewal. Many states indicated a departure from previous structures that included lifetime certification, although teachers with lifetime certification under previous structures do not have to apply for renewal certificates. Many states have also instituted mentor or entry-year programs for beginning teachers, mandating satisfaction of these requirements before full certification status is granted. In almost all states with a tiered certification structure, some type of continuing education is required. Several states, including Massachusetts and Oregon, specify the attainment of a graduate degree in order to have a teaching certificate renewed or the next level of certification awarded. Still other states offer advanced certification for teachers who have earned graduate degrees. These advanced certificates typically permit enhanced responsibilities, such as supervision and departmental administration, and are frequently valid for longer periods of time than standard certificates.

Testing requirements

In 1972, only three states implemented formal testing procedures for teacher certification (Wolfe). Erbes's 1984 study identified eighteen states that required testing for certification. Currently, seven states require no standardized test for certification. The remaining 43 states utilize a combination of basic skills and general knowledge, professional teaching knowledge, and content area knowledge tests, assessed through a variety of national and state-administered examinations, to certify teachers. Twenty-nine states employ a basic skills test. Twenty-one states administer a test of professional education knowledge. Thirty-one states require at least one test of content area knowledge (see figure 2).

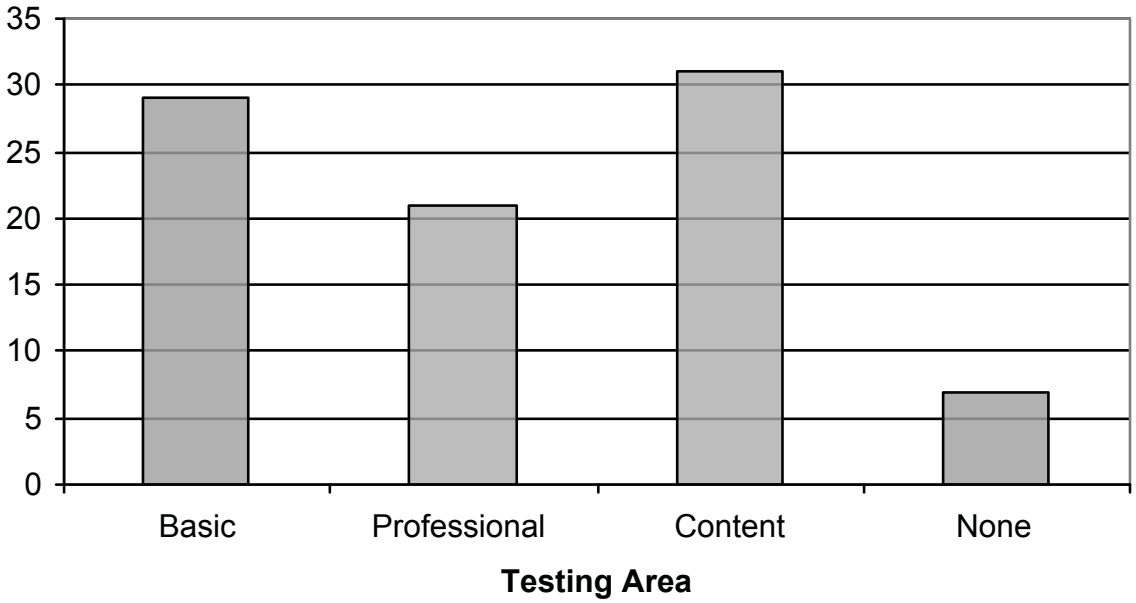

Figure 2. Types of required tests for certification.

Of these 43 states, eleven assess applicants in all three areas. Eighteen states require two tests, either in basic skills and professional knowledge (1), basic skills and content area (9), or professional knowledge and content area (8). Fourteen states use a single test in basic skills (8), professional knowledge (2), or content area (4). Of the standardized tests that are used for certification, the PRAXIS series exams (ETS, 2001) are by far the most frequent. Twenty-one states report using either the Pre-Professional Skills Tests (PPST) and/or Computer-Based Test (CBT) portions of the PRAXIS I for basic skills assessment. The Professional Learning and Teaching (PLT) portion of the PRAXIS II battery is required by 15 states. One or more of the music content exams of the PRAXIS II series is employed by 21 states. There are also 15 state-administered tests, from 11 different states, that were identified in this study. Several states also have additional requirements such as coursework in state and national constitution, Native American studies, or human relations. Most of these requirements are not waived for out-of-state applicants.

The concept of assessing qualification for teaching through standardized testing is firmly established in the requirements set forth by state certification agencies. Of the seven states currently without testing requirements, three indicated plans to implement testing requirements within the next two years. While some testing is considered necessary by almost every state, the type of tests employed by these states varies greatly. The most frequently used tests are content area tests. Many states that do not require basic skills tests indicate that they rely on the colleges and universities to determine basic skill levels before admitting students into teacher preparation programs.

Reciprocity

Most states offer some level of reciprocity for individuals desiring certification who hold valid teaching credentials in other states. Seven states claim a non-restrictive or enhanced reciprocity, in which no additional qualifications are required to obtain certification with a valid out-of-state certificate. Five states acknowledge having no reciprocity of any kind. Individuals seeking certification in these states must submit their full credentials and fulfill all requirements in the new state to obtain certification. Forty-five states claim some kind of reciprocity with other states. The level of cooperation among these states is not at all similar. In its strictest sense, the reciprocity agreements of organizations such as NASDTEC apply only to the mutual acknowledgement of regionally accredited education programs from those states. It does not exempt applicants from additional requirements such as testing or specialized coursework specified by the new state. Often, states issue a temporary credential to out-of-state teachers, allowing them time to complete these requirements during this probationary period. In many of these states, experienced teachers can be exempted from testing requirements with a minimum number of service years. In most cases, out-of-state teachers are not exempted from coursework requirements. Almost all states make exceptions for national board certified teachers, offering this elite group automatic certification. State certification specialists are hesitant to specify precise reciprocity benefits without an extended review of individual credentials.

Alternate certification programs

According to the information gathered in this study, thirty-nine states currently accommodate individuals who desire teaching certification but have non-education baccalaureate degrees by providing an alternative route to certification. The structure of these programs varies greatly. Some require that all education coursework be completed prior to teaching, while others allow for certification training while employed as autonomous teachers. These programs are typically accelerated to allow completion in one or two years. Washington utilizes an internship program, in which the candidate receives an intern-level salary and is supervised much like a student teacher until the completion of the program. Eleven states do not provide any alternative routes to certification beyond completing approved traditional education programs.

Alternate certificate programs are a relatively new addition to the certification landscape. In many cases, these programs were developed to address growing teacher shortages. In some states, such as Washington and Delaware, these programs are available only in certain subject areas. Other states actively encourage individuals to consider teaching as a second career. Virginia's alternate certification program is called the Career-Switcher Alternate Route to Licensure. Many states advertise the Troops to Teachers program as a means of obtaining teacher certification.

Fee structures

Fees assessed during the application process varied greatly among states. North Dakota assesses a $25.00 fee to obtain application materials, and is the only state to do so. Some states assess fees for evaluation of materials. Others charge an application fee regardless of the success

of the applicant in obtaining certification. Most states' fees are for the actual certification document. While the majority of these states assess a flat fee for the certificate, some base their fee on the number of certification areas or grade levels requested. Others charge by the number of years that certification will be granted. Fees for initial certification range from zero to $175.00. Four states charge no fee for certification services. Three additional states charge no fees for in-state applicants, while out-of-state applicants are assessed $10.00-$25.00. Nine other states charge a different fee for in-state and out-of-state applicants. The fee amounts identified are those assessed to initial in-state applicants. Fees for certification renewal also vary greatly. Many states have identical charges for initial and renewal certificates, while others decrease the amount required for renewal. A few states increase the amount of renewal certificates, though the number of years of certification typically increases as well. Sixteen states require a fee for fingerprinting and background checks in addition to application or certificate fees. These fees range from $22.00 to $66.00.

Accessibility

Finally, accessibility to certification also varies tremendously among the states. Thirty-five states have all application materials available to download and print. Eleven states do not. Four states have on-line forms available only for renewal or supporting documents. While all states maintain certification Web sites, many are difficult to access, lack necessary information, or present information in a confusing or contradictory manner. Unfortunately, access to information by telephone is no less accommodating. The researcher spent approximately 27 hours on the phone trying to reach a live person at state certification offices to verify information.

Conclusions

Trends in music teacher certification detected by Wolfe (1972) and Erbes (1984, 1987) have continued into the 21st century. Almost 40% of states offer multiple age-level certification; more than 66% of states have a tiered system for recertification; forty-three states require some form of testing for certification. Questions regarding reciprocity, alternate certification programs, testing fees, and on line availability of information highlighted additional facets of the teacher certification process.

State certification practices are as varied as the 50 states themselves. Specificity of age-level and content area is dependent upon individual state's needs for flexibility or "matching" desirability between teacher and classroom. More than two-thirds of the states implement a tiered certification structure, in which teachers advance through levels of certification with added experience and continuing education. To encourage continuing education, most states have abandoned lifetime certificates. Testing requirements include basic skills, professional knowledge, and content area assessments, although only 11 states assess applicants in all three areas. Eleven states have developed their own testing instruments, while the remainder employs the PRAXIS series, developed by Educational Testing Services. Most states acknowledge some level of reciprocity with other states, officially extending only to approved teacher preparation programs. Alternate routes to certification are available in approximately three-fourths of the states. Rationales for instituting these programs include addressing teacher shortages and allowing career professionals to enter the classroom as a second career. Fees for certification also vary greatly. Certification charges range from zero to $175.00, with up to $66.00 in additional fees for fingerprinting in a limited number of states. Access to information is as varied as the information itself. Although increasing available on line, some information is not immediately accessible or downloadable. Certification agencies are overwhelmed with inquiries

and are not always available to respond to questions over the phone. Several states have interactive web pages or email addresses posted, which encourage electronic submission of questions. University or college education departments should be considered as viable options for obtaining certification information.

Like similar studies before it, information contained in the study will become outdated as states continue to refine certification policies and to consider other models for certification. Future research should include a periodic revisitation of state certification practices, in an effort to detect policy trends within individual states as well as overall certification trends.

Because of the changing nature of certification standards, it is imperative for those involved with music education certification in any context to commit to an occasional review of current certification practices in their home state. Knowledge of avenues for investigation of other states' certification practices is also important, although not always of immediate significance. An understanding of issues involved in certifying teachers can provide insight into individual choices in educational preparation, the development of teacher preparation program curricula, and potential certification models for future consideration by state credentialing agencies. By providing access to this information and highlighting relevant issues in certification standards, it is hoped that music educators will not be intimidated by the task of certification, but will be encouraged to take ownership of the certification process as a result of greater understanding.

References

Berry, B. (2001). No shortcuts to preparing good teachers. *Educational Leadership, 58*(8), 32-36.

Boros, M., Duling, E., Fong, B., Haas, J., Klesitz, B., & Snyder, M. (1984). Double majoring: Do or don't? *Music Educators Journal, 71*(1), 53-55.

Bruner, J. (1977). *The process of education*. Cambridge, MA: Harvard University Press.

Carnegie Task Force on Teaching as a Profession (1986). *A nation prepared: Teachers for the 21st century*. Hyattsville, MD: Carnegie Forum on Education and the Economy.

Consortium of National Arts Education Associations (1994). *National standards for arts education: What every young American should know and be able to do in the arts*. Reston, VA: Music Educators National Conference.

Educational Testing Service. (2001). *PRAXIS series: Professional assessments for beginning teachers*. Princeton, NJ: Author.

Erbes, R. (1987). A new era in teacher certification. *Music Educators Journal, 73*(6), 42-46.

Erbes, R. (1984). Entrance into the profession: The revolution in teacher certification. *Music Educators Journal, 71*(3), 34-39.

Erbes, R. (1983). *Certification Practices and Trends in Music Teacher Education*. Reston, VA: Music Educators National Conference.

Etheridge, E. (2000-2001). Alternative certification: A threat to quality. *Childhood Education, 77*(2), 94K.

Finn, C., Jr. & Madigan, K. Removing the barriers for teacher candidates. *Educational Leadership, 58*(8), 29-31, 40.

Gallegos, A. (1978). A call for universal accreditation. *Journal of Teacher Education, 29*, 24-27.

Holmes Group. (1986). *Tomorrow's teachers: A Report of the Holmes Group*. East Lansing, MI: Author.

Rowls, M. & Hanes, M. (1982). Teacher recertification: The shift toward local control and governance. *Journal of Teacher Education, 33*(4), 24-28.

Weible, T. & Dumas, W. (1982). Secondary teacher certification standards in fifty states. *Journal of Teacher Education, 33*(4), 22-23.

Wolfe, I. (1972). *State Certification of Music Teachers*. Washington D.C.: Music Educators National Conference.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

Literacy Through Literature: The Acquisition of Music Notation and Terminology Knowledge in the Choral Rehearsal

Michele Henry and Barbara Brittain
Baylor University

Musical literacy is a long-standing goal of music education. The ability to read and respond to musical notation and terminology is among the primary reasons for music education at the elementary and secondary levels. In 1994, the National Standards for Arts Education (CNAEA) identified nine standards for music. Among these goals are "reading and notating music" and "listening to, analyzing, and describing music" (p. 61). More specifically, students in grades 9-12 are considered "proficient" if they can "demonstrate the ability to read an instrumental or vocal score of up to four staves by describing how the elements of music are used," and if they can "demonstrate extensive knowledge of the technical vocabulary of music" (p. 61).

The Texas Essential Knowledge and Skills (TEKS), adopted by the state of Texas in 1998, mirror the national standards concerning music literacy stating, "The student is expected to . . . define concepts of intervals, music notation, chord structure, rhythm/meter, and musical performances using standard terminology" (117.60.1B).

Prior to the adoption of these national and state standards, other efforts identified similar goals for music education instruction, including the comprehensive musicianship movement beginning in the 1970s. Austin (1998) compared the goals of the National Standards with those of two representatives from the comprehensive musicianship movement, Symposium on the Evaluation of Comprehensive Musicianship (Boyle and Radocy, 1973) and *Blueprint for Band* (Garofalo, 1983). Within this comparison, parallels can be drawn between the National Standards of "reading and notating music" and "listening to, analyzing, and describing music" and the comprehensive musicianship models requiring students to "demonstrate literacy in conventional notation and terminology" (Boyle and Radocy, 1973) and "translative skills including music reading, score reading, and sight reading" (Garofalo, 1983).

Forbes (2001) found that "musical elements that could be taught through the work" ranked among the top seven reasons for choral directors selecting a classical composition for performance by their choir, although this was not the case for popular musical repertoire (p. 111). Apfelstadt (2000) cautioned that when selecting repertoire, "our students deserve more

than mere entertainment" (p. 19). She charged that "through the repertoire we choose, we not only teach curricular content to our students, but we also convey our philosophy in terms of what we believe students need to learn to achieve musical growth" (p. 19).

Sherburn (1984) found that high school band students taught using comprehensive musicianship principles, including a lab experience with theory, conducting, composing, and chamber performance, scored significantly higher on a test of aural and notational skills than students in the same band rehearsals but without the lab experience. Likewise, Garofalo and Whaley (1979) found that high school band students given historical notes, concept lists, and glossaries of terms during rehearsals scored significantly higher on conceptual knowledge and aural identification tests than members of a control group whose rehearsals focused on performance preparation, while also achieving a greater performance rating than the control group. Culbert (1974) found that high school band students who devoted one of every four days of rehearsal to learning musical elements described in the band literature scored significantly higher than a control group on standardized tests of music achievement.

In the choral realm, Whitlock (1981) provided supplementary materials concerning musical elements and terminology to six high school choirs in Texas. Testing in September and April revealed significant improvements in student scores. In addition, all participating choirs received superior ratings at contest that year.

Beyond the aforementioned studies, research measuring students' knowledge of musical notation and terminology consists of tests requiring students to demonstrate musical literacy through performance tasks such as sight reading. Sight-singing studies in particular have revealed minimal competency by individual students. Henry (in press) found a mean score of 10.7 out of 28, or approximately 33%, for pitch reading when testing individual students in five Arizona high choral programs (N = 183). Henry and Demorest (1994) and Demorest and May (1995) found an approximately 66% success rate for pitch and rhythm sight singing. The latter two studies focused only on students from the top choirs in schools known for their sight-singing excellence in the state of Texas, as evidenced by success at group sight-reading contest.

Despite the generally poor showing reflected in these studies, it is likely that musical learning has occurred that was not measured, possibly that which is more rudimentary than the complex task of sight singing. The knowledge of music notation and terminology is a prerequisite for sight-reading skill development. Research focusing on playing or singing skills based on notation presupposes knowledge of the notation symbols. Perhaps students have this knowledge but have yet to apply it to performance tasks. While the acquisition of sheer factual information may seem a significantly lower level task, and a less musical one at that, it is also an appropriate starting point for inquiry into musical literacy.

Because most secondary-level music instruction takes the form of performance ensembles, it is possible that non-performance goals such as those mentioned above may receive less attention when balanced against the performing demands of most secondary school ensembles. In order to discover whether the various non-performance national and state standards for music education are being achieved through performance instruction, data must be gathered and analyzed. The Music Educators National Conference (1996b) identified recommended assessment strategies that correlate with the *National Standards for Music Education* (CNAEA, 1994). Recommended strategies for assessing achievement standard 6B, "Students demonstrate extensive knowledge of the technical vocabulary of music," include asking students "to explain and identify, orally or in writing, [music] terms or symbols" (p. 111).

Method

The purpose of this study was to determine the musical notation and terminology knowledge of students in a choral ensemble, based on the rehearsal and performance of choral repertoire from the musical score. The current study sought to determine the music literacy proficiency level by having them explain and identify in writing music terms and symbols. Research questions for this study included:

a) To what extent do students participating in high school choral programs understand music terminology and notation contained in specific performance repertoire?

b) Are other factors, including private voice study, years of choral experience, instrumental experience, keyboard experience, theory instruction, gender, and age, related to proficiency in identifying music notation and terminology?

For the purposes of this study, musical literacy was defined as the ability to understand and define written music notation, symbols, and terminology. The body of music notation, symbols, and terminology used in the study came directly from the music that students have studied and encompassed those items identified in *Performance Standards for Music* (1996b) as basic level terminology. Symbols commonly used in music, along with note values and pitch names on the staff, constituted the information included on the survey document.

Prior to testing, the researchers acquired copies of the octavo that the choir was learning for an upcoming concert. A 30-item quiz was constructed from music notation symbols and terminology within these octavos. The choir rehearsed the music for approximately seven weeks and performed these pieces on a concert. Testing occurred approximately two weeks after the concert.

Study participants were members of a non-auditioned women's choir in a Texas 5A high school (N = 57). Each participant completed a questionnaire concerning their musical backgrounds and the 30-item quiz about music terminology and notation symbols. The choir director also completed the quiz. The teacher's responses were collected for use in the event that students defined a term or symbol in a non-standard way but in a manner consistent with instruction by the director.

Results

A total of 57 students completed the questionnaire and quiz. The first section of the quiz contained 18 questions and asked students to identify the proper musical names for specific music symbols (e.g. treble clef, sharp). The second section requested that students define as specifically as possible music terms (e.g. *legato*, *rit*.). This section contained eight questions. The final section asked the students to provide the note letter name for four different pitches. Because the population was a women's choir, all pitches were located on the treble clef. All material for the quiz came directly from the scores of the pieces studied in choir.

The overall mean score was 10.44 out of 30, or approximately 33%, with a standard deviation of 6.96. The mean score for section one was 8 out of 18, or approximately 44%. The mean score for section two was 1.22 out of 8, or approximately 15%. The mean score for section three was 1.21 out of 4, or approximately 30% (see figure 1).

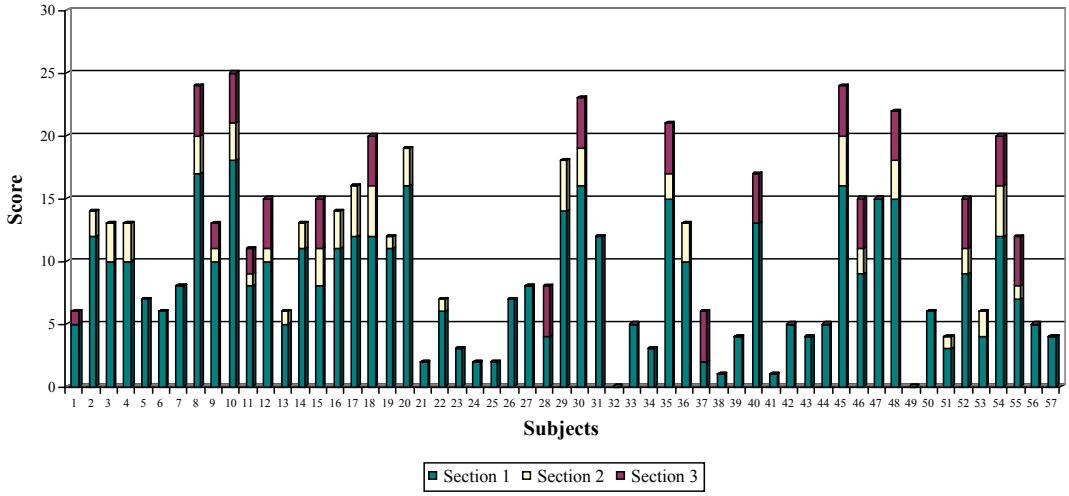

Figure 1. Scores by subject divided by section (n = 57). Each individual bar represents the total score for each subject. Scores in black indicate the number of correct responses for the notation section, out of a possible score of 18. Scores in white indicate the number of correct responses for the terminology section, out of a possible score of 8. This score is added to the notation score. Scores in gray indicate the number of correct responses for the note naming section, out of a possible score of 4. This score is added to the previous two sections.

The questionnaire requested background information including grade, voice part, years of choral experience, private voice experience, music theory study, keyboard experience, and instrumental experience. A multiple regression was run using these variables as well as total score. These predictor variables accounted for 35% of the variance on the scores. Individually, choral experience and keyboard experience were statistically significant at the .05 level.

Discussion

Subjects demonstrated greatest success in the first section of the quiz (44% accuracy). The names of the musical symbols were the most familiar to the students. Because many of these symbols remain constant within most musical scores, the frequency with which students encounter them is most likely the highest as well. The second section of the quiz revealed the lowest achievement (15%). Because specific musical terms such as *legato* occur less frequently, students will have less recurring exposure to them.

The third section of the test presented some logistical problems that were not discovered until scoring occurred. Because the school utilizes solfège when teaching music reading, students were not necessarily taught the letter names of the notes on the staff. Many students did not understand the difference between letter names on the staff and solfège letter names. Most students responded by identifying solfège syllables. The fact that no key had been specified further complicated the issue. Because there was no key signature, it could be presumed that all syllables should be given for the key of C. However, most students seemed to randomly select a *do* for the set of questions (most often F). When scoring occurred, students were given credit only if the actual letter name of the note was correctly identified. Although the results for this section are higher than for section two of the quiz (30% accuracy), these results are somewhat

misleading. The few students who correctly understood the task tended to answer all four questions correctly, while those who did not understand the directions scored zero. So while the results could be interpreted that subjects knew approximately one of four questions in this section, it is likely more accurate to state that approximately one out of four students understood the question. As a result of this confusion, and the fact that many choral directors focus on solfège syllables to the exclusion of note letter names, this section of the quiz will be dropped from future testing.

These initial results are promising for choral music educators, insofar as the significant relationship between choral experience and knowledge of music notation and terminology, in particular, indicates that music literacy learning is occurring within the choral classroom. The significant correlation between keyboard experience and success on the quiz is consistent with previous findings that keyboard experience is related to success on other measures of music literacy (Cooper, 1954; Demorest, 1996; Demorest and May, 1995; Henry and Demorest, 1994; Thostenson, 1969).

Because this was a preliminary study, no effort was made to ascertain students' knowledge of musical terminology and symbols prior to instruction on the choral literature taught during the seven-week period. While the overall scores are not strong, it should be noted that the subjects were members of a beginning choir and had relatively little formal music training. Over half of the subjects reported three years or less choral experience. In subsequent investigations, a pre-test posttest design will be employed to quantify growth in this type of musical knowledge. Effort will need to be given to minimize test effect in this situation. Additionally, larger test populations at multiple schools will be employed to allow greater generalization of the findings by accounting for teacher effect, socio-economic status, and other factors likely to remain constant within one school population. Future research efforts could also explore this notation and terminology learning in the band and orchestra classrooms, where performance is typically more dependent upon music reading knowledge.

References

Apfelstadt, H. (2000). First things first: Selecting repertoire. *Music Educators Journal*, *87*(1), 19-22.

Austin, J. (1998). Comprehensive musicianship research: Implications for addressing the national standards in the music ensemble classes. *UPDATE: Applications of Research in Music Education, 17*(1). 25-32.

Boyle D. & Radocy, R. (1973). Evaluation of instructional objectives in comprehensive musicianship. *Bulletin of the Council for Research in Music Education*, *32*, 2-21.

Consortium of National Arts Education Associations (1994). *National standards for arts education: What every young American should know and be able to do in the arts.* Reston, VA: Music Educators National Conference.

Cooper, J. (1954). The development of a sight-singing achievement test for use with college students. Unpublished doctoral dissertation, University of Colorado, Boulder.

Culbert, M. (1974). The effects of using a portion of the rehearsal time for developing skills in describing music on the performance level and musical achievement of high school band students. Temple University, 1974. *Dissertation Abstracts International, 35,* 3486A.

Demorest, S. (1998). Improving sight-singing performance in the choral ensemble: The effect of individual testing. *Journal of Research in Music Education*, *46*, 182-192.

Demorest, S. & May, W. (1995). Sight-singing instruction in the choral ensemble: Factors related to individual performance. *Journal of Research in Music Education, 43,* 156-167.

Forbes, G. (2001). The repertoire selection practices of high school choral directors. *Journal of Research in Music Education*, *49*(2), 102-121.

Garofalo, R, (1983). *Blueprint for Band*. (Rev. ed.). Ft. Lauderdale, FL: Meredith Music Publications.

Garofalo R. & Whaley, G. (1979). Comparison of the unit study and traditional approaches for teaching music through school band performance. *Journal of Research in Music Education*, *27*, 137-142.

Henry, M. (in press). The development of a vocal sight-reading inventory. *Bulletin of the Council for Research in Music Education.*

Henry M. & Demorest, S. (1994). Individual sight-singing achievement in successful choral ensembles. *UPDATE: Applications of Research in Music Education, 13,* 4-8.

Music Educators National Conference (1996). *Performance standards for music: Strategies and benchmarks for assessing progress toward the national standards*. Reston, VA: Author.

Sherburn, E. (1984). Student achievement and attitude in high school instrumental music education: A comparison of the effects of a lab approach and a more traditional approach. University of Southern California, 1984. *Dissertation Abstracts International, 45,* 804A.

Texas Education Agency (1998). *Texas Essential Knowledge and Skills*. Austin, TX: Author.

Thostenson, M. (1969). The study and evaluation of certain problems in ear training related to achievement in sight singing and music dictation. *Bulletin of the Council for Research in Music Education, 11,* 14-35.

Whitlock, R. (1982). The design and evaluation of study materials for integrating musical information into the choral rehearsal. North Texas State University, 1981. *Dissertation Abstracts International, 42,* 4643A.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

Musical Aptitude and Student Achievement of Beginning Instrumental Music Students: Results of Second-Year Study

J. Richard Holsomback, Jr.
Pine Tree Independent School District, Longview, Texas

The decade of the 1990's for public education in our nation could be characterized as a decade of increased emphasis upon school accountability. The accountability standard has been results that individual school districts achieve with regard to progress made on mandated standardized tests. Although internal pressure for increased student achievement as measured by these tests are evident in our schools, external pressure from the business community and the public at large have played a major role in the school accountability movement throughout the nation during the decade of the 1990's.

In their efforts to meet the challenges of the emphasis upon standardized testing, music educators have been asking themselves: What knowledge is worth knowing in music and what role will music play in the context of a diverse society in the twenty-first century? What should music programs look like with regard to organization for learning? What benchmarks should be used in predicting student success in instrumental music programs? What relationships exist between musical aptitude or musical achievement and academic achievement? Before music educators can answer these questions, students must be placed into the instrumental program. Frequently, one of the benchmarks used for placement has been a musical aptitude assessment. The measurement and evaluation of such a construct as musical aptitude has become a particular focus in the continuously evolving area of the psychology of music and such assessments are also of practical significance to music educators.

According to Boyle (1992), the area of assessment of musical constructs can be traced to the work of Carl Seashore. Seashore's (1919) early efforts to assess music talents provided the model and the impetus for the psychology of music research in the first half of the twentieth century. During the later part of the century, three theories surfaced regarding the nature, development and evaluation of various musical constructs, especially the construct of musical aptitude. Probably the most prominent were those of Gordon (1987) and his theory of

developmental musical aptitude, Gardner's (1983) theory that musical intelligence was one of several loosely related multiple intelligence and the work of Karma (1985) who postulated the view that musical aptitude involved perceptual/cognitive structuring of acoustical material. Some of the early testing of these theories and testing the relationship of musical aptitude and its relationship to student achievement are evident in the literature.

For example, Young (1971) used the *Musical Aptitude Profile* of Gordon (1965), the *Lorge-Thorndike Intelligence Test* (Lorge and Thorndike, 1962) and the *Iowa Test of Basic Skills* (1962) to predict success in an elementary school instrumental program in suburban Chicago. Young concluded that student success in instrumental music was best predicted by the use of the Musical Aptitude Profile, the intelligence test and the student achievement test. Young also asserted that the relationship of musical aptitude to musical achievement increased with years of study, while the relationship of intelligence and academic achievement to musical achievement, while initially higher, became noticeably less with additional years of instrumental music study.

Most of the other literature on musical aptitude has its focus upon the effect of participating in musical activities in school and the possible link to that participation upon students' standardized achievement scores. In addition, according to McIntyre and Cowell (1984), most of the results of these studies are unclear and are often contradictory. For example, Kvet (1985) found no significant difference in sixth-grade reading, language and mathematics achievement between students who were excused from regular classroom activities for the study of instrumental music and students not studying instrumental music. Camp (1990), on the other hand, found that participation in extra-curricular and co-curricular activities such as instrumental music had a positive relationship to grades. Similarly, more recent research has determined a significant relationship between musical sound discrimination and reading ability of first grade children.

The evidence comes from a recent study of Lamb and Gregory (1993). In addition to some standard reading tests, children were tested on their ability to "sound out" nonsense syllables that they viewed on cards (phonic reading) and pitch awareness, in which they heard pairs of musical notes or chores in sequence and reported whether they sounded the same or different, much like pitch discrimination in musical aptitude profiles. Also the children were tested with notes that had the same or different timbres. Their phonic awareness was assessed by listening to spoken words and telling whether the words began or ended with the same sound. They found a high degree of correlation between how well children could read both standard and phonic material and how well they could discriminate pitch.

Method

The purpose of this study was to test the relationship between musical aptitude and academic achievement of instrumental students across two academic years. Academic achievement in this study was defined as scores on the Texas Assessment of Academic Skills Reading and Mathematics Tests (TAAS). In addition, all sub-tests of the reading and mathematics tests were also examined. Musical aptitude was defined as composite scores of the Selmer Music Guidance Survey. The following questions were addressed descriptively and analytically: Did a relationship exist between musical aptitude and the Texas Assessment of Academic Skills Test? If such a relationship existed, did the relationship exist only in composite scores of the reading and mathematics test or did the relationships also exist between musical aptitude and sub-test of the TAAS reading and mathematics tests? The hypothesis stated that no relationship existed

between musical aptitude and the Texas Assessment of Academic Skills Test. Significance was set at the .05 level ($p<.05$).

For this study it was assumed that comparable data were available on the number of pupils in the district under study and that data on the variables, as defined in this study, were available for all students being studied in a standardized format.

Limitations of this study included:

1. This study was limited to one East Texas independent school district and the study was limited to two academic years. The study traced the beginner students and their scores at the beginner level and the next level of their instrumental music program.

2. All data were ascertained through data sent to the school district by the Texas Education Agency. Definitions of the variables, especially the variables that defined student achievement, were important because other variables and school site data could yield very different results.

This basic design was a correlation study used to determine the relationship between musical aptitude and acadimi9c achievement of beginner band students as measure by standardized tests used in this particular study. The sampling technique was a purposive technique. The sample consisted of 104 sixth grade band students in an East Texas school district for the pilot study (Holsomback, 2001) and 74 of those students were available for study in the second-year data sets. According to Gay (1981) 30 subjects are considered to be a minimally acceptable sample size for a correlation study. This sample met and exceeded that particular benchmark in the literature. Since no research was found in the literature with regard to the Selmer Music Guidance Survey, the reliability coefficient was established for the composite scores. A split-half procedure was employed to establish the reliability coefficient on the Selmer Music Guidance Survey composite scores. Likewise, the Spearman-Brown prophecy formula was utilized for correction purposes.

Students were given the Selmer Music Guidance Survey to serve as a guide to instrument assignment and general assessment of the individual needs of the selected band students at the beginning of the school year. Students are then interviewed for physical characteristics to match those characteristics for instrument placement in the sixth grade year, the year in which these students began their instrumental music training. Therefore, the sample was an intact group for research.

To measure student achievement, the Texas Assessment of Academic Skills Test was given to the students in April 2000 for the initial pilot study and the tests were administered again in April 2001. The data was then ascertained for the second-year study for the 74 students that remained in the band program for their seventh grade year. The data sets were then complete across two academic years. Similarly, all students were given these tests as part of the regular testing program for the entire school district as well as the band students sampled in this study.

Using Liebetrau (1983) as a guide, the Pearson product-moment correlation was computed, examining the relationship between the Selmer Music Guidance Survey and the Texas Assessment of Academic Skills Test. The hypothesis was rejected and the research question was answered for the standardized assessments used in the study if the correlation result met or exceeded the critical value at the .05 level ($p<.05$).

Results

The descriptive analysis yielded a mean on the Selmer composite scores for the first academic year (1999-2000) of 85.038 and a standard deviation of 9.038. The coefficient of variation across the entire distribution was .100 or a 10% variation of the Selmer composite scores for the entire distribution for the composite score for the first academic year. For the second-year data sets (2000-2001) the sample size was 74 due to various attrition factors and that sample yielded a Selmer composite score of 85.742 and a standard deviation of 8.844. The coefficient of variation across the sample was .103 or a 10.3% variation across the entire distribution of composite scores.

Once the descriptive statistics were ascertained for the data, a reliability coefficient analysis was performed on the Selmer Music Guidance Survey for the composite scores. The split-half method was chosen to determine reliability. The entire Selmer Music Guidance Survey was administered to the entire sample of student used in this study. The 60 items for the composite tests were divided into odd and even sets of data. The Spearman-Brown prophecy formula was employed to correct the reliability coefficient for the Selmer Music Guidance Survey composite scores. The formula yielded a reliability coefficient of .84 for the Selmer composite scores. Likewise, the standard error of measurement was 1.4. The reliability coefficient for the Selmer composite scores compared to a split-half reliability coefficient of .80 and a standard error of 1.5 for the 80-item Intermediate Measures of Music Audiation by Gordon (1986).

Once the descriptive statistics and the reliability coefficient had been established, the Pearson Product Moment correlation was computed for the Selmer Music Guidance Survey composite against the Texas Assessment of Academics Skills Test (reading and math) composite scores. The correlation for the TAAS reading composite and the Selmer Music Guidance Survey composite was .494 for the 104-student sample. The correlation coefficient for the second-year 74 student sample was .360. Both correlation coefficients between the Selmer Music Guidance Survey composite scores and the TAAWS reading composite scores were positive and significant at the .05 level ($p<.05$) across two academic years.

Once the composite reading scores were examined, the analysis progressed to the Texas Assessment of Academic Skills Reading sub-tests. The sub-tests for TAAS Reading included word meaning in a variety of written formats, recognition of supporting ideas, summarization technique of reading passages and recognition of relationships and outcomes. In addition, recognition of inferences and generalizations were included in the TAAS Reading sub-tests along with recognition of points of view, propaganda and fact and opinion. The correlation for the TAAS reading word meaning sub-test and the Selmer Music Guidance Survey composite was .233 for the 104-student first-year sample. The correlation coefficient for the second-year 74-student sample was .187. Both correlation coefficients between the Selmer Music Guidance Survey composite scores and the TAAS reading word meaning sub-test scores were positive but not significant at the .05 level ($p<.05$) across two academic years.

The correlation for the TAAS reading supporting ideas sub-test and the Selmer Music Guidance Survey composite scores was .232 for the 104-student first-year sample. The correlation coefficient for the second-year 74-student sample was .109. Both correlation coefficients between the Selmer Music Guidance Survey composite scores and the TAAS reading word meaning sub-test scores were positive but not significant at the .05 level ($p<.05$) across two academic years. The summarization sub-test correlation to the Selmer was .446 for

the first-year sample and .332 second-year sample. Both correlations between summarization years employed in the study. The correlations for relationships and outcomes to the Selmer Music Guidance Survey were .295 and .567 respectively across two academic years. Likewise, the correlations for inferences and generalizations were .453 and .258 across the two years of study and the correlations for point of view , propaganda and fact and opinion were .399 and .136 respectively across two academic years used in the study. Both correlations for relationships and outcomes were positive and significant at the .05 level for two academic years. This was also true for inferences and generalizations. However, for points of view, propaganda and fact and opinion, the correlation was positive and significant for the first academic year but not the second academic year. The results for the TAAS Reading composite scores and all sub-tests are presented in Table 1.

Once the analysis was completed for the TAAS Reading tests, the results were ascertained for the Texas Assessment of Academic Skills Mathematics Test (TAAS) and the Selmer Music guidance Survey composite scores. The correlation coefficients for the TAAS mathematics composite scores and the Selmer were .378 for year one and .263 for year two. The correlations were positive and significant at the .05 level ($p<.05$) across both years. The correlations for number concepts and the Selmer were .242 for year one and -.025 for year two. For algebraic concepts/mathematical relations and functions the correlations were .218 and .160 respectively for years one and two. None of the correlations were significant at the .05 level across two years of data. The results for the mathematics composite scores and all subtests are presented in Table 1.

The correlations for geometric properties and relationships and the Selmer were .299 for year one and .190 for year two. For measurement concepts the correlations were .384 and .221 respectively for years one and two. Likewise, the correlations for probability and statistics were .361 for year one and .033 for year two. The correlations for these sub-tests were significant at the .05 level for these same sub-tests for year two.

In the next mathematics sub-tests, the use of addition, subtractions, multiplication, and division operations are used to solve problems. The correlations for addition were .281 for year one and .319 for year two. Both correlations were significant at the .05 level ($p<.05$). For subtractions the correlations were .138 for year one and .012 for year two. None of the correlations were significant at the .05 level for two years of data sets. For multiplication sub-test scores and the Selmer Music Guidance Survey composite scores the correlations were .132 for year one and -.095 for year two. Both correlations were not significant at the .05 level. For the use of division to solve problems the correlations to the Selmer composite were .250 for year one and .272 for year two. Both correlations were significant at the .05 level for the use of division to solve problems on the TAAS Mathematics test.

In the final mathematics sub-tests, estimations, solution strategies, mathematical representations and evaluations of the reasonableness of a solution are used in problem solving. For problem solving using estimation the correlations to the Selmer Music Guidance Survey composites scores were .316 for year one and .064 for year two. Problem solving using solutions strategies yielded correlations of .175 for year one and .260 for year two. The correlation for problem solving using estimation was significant at the .05 level ($p<.05$) for year one but it was not significant at the .05 level in year two. The correlation for problem solving using solution strategies was not significant for year one but was significant at the .05 level for year two.

Table 1
Product Moment Correlation Coefficients for the Texas Assessment of Academic Skills Reading and Mathematics Sub-Tests Scores and the Selmer Music Guidance Survey Composite Scores

Texas Assessment of Academic Skills Test	Selmer Composite Year 1	Year 2
Reading Composite Scores	0.494	0.360
Reading Sub-Test Scores		
Word Meaning	0.233	0.187
Supporting Ideas	0.232	0.109
Summarization	0.446*	0.332*
Relationships and Outcomes	0.295*	0.567*
Inferences and Generalizations	0.453*	0.258*
Point of view, propaganda and fact and opinion	0.399*	0.136
Mathematics Composite Scores	0.378*	0.263*
Mathematics Sub-Test Scores		
Number Concepts	0.242	-0.025
Algebraic/Mathematical Relations and Functions	0.218	0.160
Geometric Properties and Relationships	0.299*	0.190
Measurement Concepts	0.384*	0.221
Probability and Statistics	0.361*	0.033
Use of Addition to Solve Problems	0.281*	0.319*
Use of Subtraction to Solve Problems	0.138	0.012
Use of Multiplication to Solve Problems	0.132	-0.095
Use of Division to Solve Problems	0.250*	0.272*
Problem Solving Using Estimation	0.316*	0.064
Problem Solving Using Solution Strategies	0.175	0.260*
Problem Solving Using Mathematical Representation	0.380*	0.348*
Evaluation of the Reasonableness of a Solution	0.266*	0.332*

Source: Texas Assessment of Academic Skills Reading and Mathematics Tests (Spring 2000, Spring 2001) and the Selmer Music Guidance Survey Composite Scores.
Note: *$p<.05$

Conclusions

The purpose of this study was to test the relationship between musical aptitude and academic achievement of instrumental music students across two academic years. Achievement in this

study was defined as scores on the Texas Assessment of Academic Skills Reading and Mathematics Tests (TAAS) and all of its subtests. Musical aptitude was defined as composite scores on the Selmer Music guidance Survey. The results indicated that a relationship did exist between musical aptitude and the TAAS scores. The relationships for composite reading and composite mathematics were positive and significant at the .05 level across two academic years. In addition, in the reading subtest areas of summarization, relationships and outcomes and inferences and generalizations, the relationships were positive and significant at the .05 level. Similarly, the use of addition and division to solve problems, problem solving using mathematical representation, and evaluation of the reasonableness of a solution mathematics subtests were all positive and significant at the .05 level across two academic years. The null hypothesis was rejected. It stated that no relationship existed between musical aptitude and student achievement as measured by the achievement tests used in this study. These positive and significant relationships also hold up across two academic years; therefore, more research needs to be conducted to see if these results are consistent across even longer periods of time. The implications for instrumental music educators are:

1. Musical aptitude tests should be combined with other benchmarks, such as teacher recommendations, performance on standardized achievement tests, student instrument preference, and instrumentation needs for student selection into the program.
2. The Selmer Music Guidance Survey composite scores are at least equally reliable as the Gordon Intermediate Measures of Musical Aptitude on a split-half procedure, and
3. The study suggests that more research needs to be conducted since the findings are in agreement with the research of such studies as those of Lamb and Gregory (1993), which concluded that a high degree of correlation existed between how well children read both standard and phonic material and how well they discriminate pitch. It should be noted that the Selmer Survey also included a pitch component as part of its assessment of musical aptitude.

The findings of this study do nothing to solve the theoretical aspects of musical aptitude. Questions still exist as to whether musical aptitude is developmental, integrated into other aspects of intelligence in children or whether musical aptitude involves more cognitive structuring of acoustical materials. Even in light of the fact that these important theoretical questions remain, it is clear from the findings that musical aptitude has important relationships to academic achievement. Educators and administrators alike now have compelling evidence that music is a valuable link in our modern curriculum. It is not clear as to whether music can actually make a child “smarter” but we can now see that there is a vital link between a child’s musical aptitude and their ability to perform important reading and mathematical tasks. These tasks also appear to be not just minimal tasks but tasks that have long been considered higher order thinking skills as well.

References

Boyle, J.D. (1992). Evaluation of music ability, In Colwell, R. (Ed.) *Handbook of research on music teaching and learning*. New York, NY, Schirmer Books.

Camp, W.G. (1990). Participation in student activities and achievement: A covariate structural analysis. *Journal of Educational Research, 83* (5), 272-278.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Gay, L.R. (1981). *Educational research: Competencies for analysis and application*. Columbus, OH: Merrill.

Gordon, E. (1965). *Musical Aptitude Profile*. Boston, MA: Houghton Mifflin.

Gordon, E. (1986). *The intermediate measures of music audiation*. Chicago, IL: G.I.A. Publications.

Gordon, E. (1987). *The nature, description, measurement, and evaluation of music aptitude*. Chicago, IL: G.I.A. Publications.

Holsomback, R. (February 2001). *Evaluating the relationship between musical aptitude and standardized achievement test scores of beginning instrumental music students*. Poster session presented at the annual meeting of the Texas Music Educators Association, San Antonio, TX.

Iowa Test of Basic Skills (1962). Boston, MA: Houghton Mifflin.

Karma, K. (1985). Components of auditive structuring: Towards a theory of musical aptitude. *Bulletin of the Council for Research in Music Education, 82*, 1-13.

Kvet, E. (1985). Excusing elementary school students from regular classroom activities for the study of instrumental music: The effect of sixth-grade reading, language, and mathematics achievement. *Journal of Research in Music Education, 33* (1), 45-54.

Lamb, S. and Gregory, A. (1993). The relationship between music and reading in beginning readers. *Educational Psychology, 13*, 19-26.

Liebetrau, A. (1983). *Measures of association*. Thousand Oaks, CA: Sage Publications.

Lorge, I. and Thorndike, R. (1962). *Lorge-Thorndike Intelligence Test*. Boston, MA: Houghton Mifflin.

McIntyre, T. and Cowell, K. (1984). *The use of music and its effects on the behavior and academic performance of special students: A review of literature*. Ed. 332447.

Seashore, C. E. (1919). *The psychology of musical talent*. New York: Silver Burdett.

Young, W. (1971). The role of musical aptitude, intelligence, and academic achievement in predicting the musical attainment of elementary instrumental musical students. *Journal of Research in Music Education, 19* (4), 385-398.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

A Comparison of Music Teachers' and Their Students' Perceptions About Student Practice Versus How They Actually Practice

Lisa M. Maynard, Georgia A. Green, and Carl J. Poposki
Baylor University

A great deal of emphasis is placed on the importance of practice in the lives of musicians. This emphasis begins at the time a student first learns to play an instrument or sing, and continues through out their lives as musicians. While many teachers of music employ strategies such as having their students fill out practice time sheets, writing specific practice instructions in homework books, or audio taping their practice, very few are privy to seeing what their students actually do in the privacy of their practice rooms.

While some research literature exists on the topic of practice, investigations have so far mostly involved comparisons of different practice types and strategies (Britt, 1997; Coffman, 1990; Gruson, 1988; Pressing, 1988; Rosenthal, 1984; Ross, 1985; Shin, 1990; Zurcher, 1975) rather than investigating the more specific aspects of what individuals actually do when they practice (Geringer & Kostka, 1984; Madsen & Geringer, 1981; Maynard; 2000; Maynard; 2000; Miklaszewski, 1989). Less is known about the relationship between music teachers' perceptions of how their students practice versus what their students actually do in their practice. Also relevant to this area of research are the students' own perceptions of how they practice, and if these perceptions change in any way after observing their own practice. Prior research investigating the topic of instructor and student observations and impressions of teacher behaviors suggest that students tend to rate themselves higher after observing themselves teaching on video tape (Cassidy, 1993).

The purpose of this study was to investigate how music teachers' perceptions of their students' practice (before actually observing them practice) might differ from their impressions of the same students' practice skills after observing what they actually do in their practice sessions.

Method

Subjects were 5 music teachers of double bass, oboe, voice, french horn, and piano, and 5 undergraduate music students from a major University School of Music. Prior to beginning the study each music teacher was asked to recommend one student from their studio who they thought was highly proficient at practice. After final decisions were made, the teachers were asked to provide ratings for a series of 10 questions pertaining to their perceptions as to the effectiveness of their students' practice from behaviors observed in private lesson settings. A Likert scale using 1 for least effective to 5 for most effective was employed for this purpose.

The questionnaires asked subjects to rate practice efficiency, attention to detail (e.g. phrasing, dynamics, changes in tempo), intolerance for mistakes, attentiveness, rhythmic precision, attention to correct intonation/notes, metronome usage, singing or vocalizing in practice (out of context), tenacity for repetition, and the overall quality of practicing.

After accepting their teacher's offer to participate in the study, each student was contacted by the researchers to set up a time to be video taped while practicing. For the purposes of this study, students were asked to practice 30 minutes of own choice solo repertoire in a location of their choosing. Before taping began each student was asked to provide ratings for the same set of questions that their teachers had answered except the responses in this case were based on the students' self perceptions of how they thought they practiced. Shortly after the videotaping of his or her practice session had concluded, each student was asked to view his or her practice tape in its entirety. Immediately following, each student was asked once again to provide ratings for the same questions initially asked, but this time focusing their responses on what they had observed on their practice tape.

Once each student had completed both questionnaires, their music teacher viewed the entire 30 minute practice videotape of his or her student. Immediately following this the music teachers once again rated their student's practice using the questionnaire they had originally been given but this time responding to their observations of how their students actually practiced on the videotape.

From the individual data collected from the questionnaire responses, two sets of mean scores each were calculated for both the students and the music teachers: one representing each group's rating designations prior to viewing the practice tape, and the other the ratings given after viewing the tape. Table 1 illustrates these mean scores. Comparisons of these means were then made both within the student and teacher groups respectively, and between the student and teacher responses.

Table 1
Pre- and Post-Mean Scores of Teacher and Student Groups

Practice Questions	Efficiency	Detail	Mistakes	On Task	+ RH.	+ Pitch	Metro-nome	Sing/ Voc.	Reps.	Overall	Total Score
Means S Pre	*3.2*	*3.2*	*3.0*	*3.6*	*3.0*	*3.0*	*2.6*	*2.6*	*3.2*	*3.2*	*3.06*
Means	*3.6*	*3.6*	*3.0*	*4.0*	*3.4*	*3.2*	*3.0*	*2.4*	*3.6*	*3.6*	*3.34*

S Post											
Means T Pre	*3.2*	*3.4*	*3.0*	*3.8*	*3.2*	*3.4*	*3.0*	*2.6*	*3.6*	*3.4*	*3.26*
Means T Post	*3.2*	*3.0*	*3.4*	*3.6*	*3.2*	*3.6*	*2.4*	*3.2*	*3.8*	*3.0*	*3.26*

Mean data were also calculated according to instrument to see what differences, if any, existed between the groups by instrument as illustrated by Table 2.

Table 2
Pre- and Post-Mean Scores by Instrument Groups

Instrument Group	Double Bass	Oboe	Voice	French Horn	Piano	Total
Pre	4.7	8.2	7.6	5.9	5.2	6.32
Post	5.0	8.2	8.4	5.2	6.2	6.6
Differences	+.3	0	+.8	-.7	+.1	+.28

Results

Within the student subject group, the means for all of the questions asked post- video tape viewing (with the exception of one) were at least equal to or higher than the means for the ratings given on the initial questionnaire. Only 1 of the post-viewing student mean scores (singing/vocalization) was lower than the remaining 9 pre-viewing mean scores. Of these 9, 7 were .4 higher, 1 was .2 points higher, and 1 remained the same. This suggests that the student subjects' actual opinions about their practice after having watched the video tapes were higher than their initial perceptions of how they might practice but were not noticeably so. The total mean score for all 10 practice questions was greater in the student group's post-viewing results by .28 points.

Of the music teachers' pre- and post-viewing mean scores, 2 of the means for the questions related to efficiency and rhythmic precision in practice remained the same, while mean scores for detail, on task practice behavior, use of a metronome, and overall practice rating were all lower by .4, .2, .6, and .4 respectively in the post-viewing responses. The remaining post-viewing mean scores were higher in the questions related to tolerance for mistakes (.4), pitch accuracy (.2), singing/vocalization (.6), and propensity for repetition (.2). The total mean score for all 10 practice questions did not change from the music teacher group's pre-viewing score to their post-viewing score (3.26).

Comparisons of means were also made among groups of instruments. The pre- and post-mean scores were highest for the oboe (B=8.2, A= 8.2) and lowest for the double bass (B=4.7, A=5.0). The remaining instrument groups pre- and post-means were as follows: voice (B=7.6), (A=8.4); piano (B=5.2, A=6.2); and french horn (B=5.9, A=5.2). The greatest increase in mean occurred between the individual pre- and post-mean scores of the voice subjects (an increase of .8). While the largest decrease of mean score occurred between the pre- and post-mean values of the french horn (down by .7).

The overall mean score of the combined instrument groups' pre- scores was 6.32, while the overall post-mean for all groups was higher at 6.6. The numeric difference between the pre- and post-mean scores of all the instruments was .28. Using these group instrument pre- and post-

means, a comparison was made with the individual instrument groups means. In the pre- mean category the voice (7.6) and oboe (8.2) groups means were above the combined pre- mean for all instruments, while the horn (5.9), piano (5.2), and double bass (4.7) were lower.

The oboe (8.2) and voice (8.4) once again had post-mean scores higher than the overall mean score for all the instrument groups except this time in reverse. The remaining means for piano (6.2), and horn (5.2) were also in reverse order to the pre- mean score order, while the double bass (5) remained in the same order.

Discussion

The findings of this study provide possibilities for future research in assessing how musicians think about and carry out their practice routines. In training novice teachers video taping is used frequently as a means by which to improve instruction but as previous and this current research suggests, student self-ratings of performance whether relating to teaching or practice seem to have a tendency to be higher after watching themselves on tape. The results of this study found that the majority of students in the student group rated their practice higher post-video. In perhaps stark contrast to this finding, the total mean score for all 10 practice questions did not change from the music teacher group's pre-viewing score to their post-viewing score (3.26) perhaps supporting the notion that experienced music teachers perceptions of how their students might be practicing at home are somewhat accurate just by impressions formulated in private lesson situations. Further research needs to be done to ensure that students are practicing at all times in a manner that is both efficient and effective, and that will lead to the highest possible levels of music proficiency.

References

Britt, M. E. (1997). *The effects of computerized visual reinforcement on the development and enhancement of selected trombone techniques: A case study (mental imagery)*. Unpublished doctoral dissertation, Florida State University.

Cassidy, J.W. (1993). A comparison between students self-observation and instructor observation of teacher intensity behaviors. *Bulletin for the Council of Research in Music Education, 115*, 15-29.

Coffman, D. D. (1990). Effects of mental practice, physical practice, and knowledge of results on piano performance. *Journal of Research in Music Education, 38*, 187-196.

Geringer, J. M., & Kostka, M. J. (1984). An analysis of practice room behavior of college music students. *Contributions to Music Education, 11*, 24-27.

Gruson, L. M. (1988). Rehearsal skill and musical competence: Does practice make perfect? In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 91-112). New York: Oxford University Press.

Madsen, C. K., & Geringer, J. M. (1981). The effect of a distraction index on improving practice and attentiveness and musical performance. *Bulletin of the Council for Research in Music Education, 66-67*, 46-52.

Maynard, L. M. (2000). *A case study of an expert music teacher's practice on the viola*. Poster session presented at the meeting of the Texas Music Educators Association. Maynard, L.M. (2000). *The role of repetition in the practice sessions of music teachers and their students*. Unpublished doctoral dissertation. The University of Texas at Austin.

Miklaszewski, K. (1989). A case study of a pianist preparing a musical performance. *Psychology of Music, 17*, 95-109.

Pressing, J. (1988). Improvisation: Methods and models. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 129-178). New York: Oxford University Press.

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*, 265-273.

Ross, S. L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. *Journal of Research in Music Education, 33*, 221-230.

Shin, R. R. (1990). *The mirror inversion piano practice method and the mirror music of Vincent Persichetti*. Unpublished doctoral dissertation, University of Alabama.

Zurcher, W. (1975). The effect of model-supportive practice on beginning brass instrumentalists. In C. K. Madsen, R. D. Greer, & C. H. Madsen, Jr. (Eds.), *Research in music behavior: Modifying music behavior in the classroom* (pp. 131-138). New York: Teachers College Press, Columbia University.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

An Investigation of the Effects of Vocal Modeling on the Development of Singing Ability in Kindergarten Children

Diane Cummings Persellin, *Trinity University*, Linda Smith, *Woodstone Elementary School, San Antonio, Texas,* Martha Klein and Emily Taguiam, *Wetmore Elementary School, San Antonio, Texas*

Singing is the most basic music skill young children learn. It is a learned behavior – a psychomotor skill. Vocal pitch accuracy improves with age and with appropriate instruction (Goetze, Cooper, & Brown, 1990). Extensive studies have detailed factors associated with improved vocal accuracy. Investigators have found that young children echo with greater accuracy when the model's voice is female rather than male (Sims, Moore, & Kuhn, 1982; Yarbrough, Green, Benson, & Bowers, 1991; Hendley & Persellin, 1996), a child's voice rather than an adult's voice (Green, 1987), and a voice without vibrato rather than with vibrato (Yarbrough, Bowers, & Benson, 1992). Hermanson (1972) found that human models have a more positive effect on vocal accuracy than instrumental models. Kinesthetic reinforcement of a vocal model (Murphy & Persellin, 1993; Youngson & Persellin, 2001) and the use of learning modalities (Apfelstadt, 1984; Persellin, 1993; 1994) have also been shown to have a positive effect on young children's vocal accuracy.

Goetze (1986) found a neutral syllable to facilitate more accurate singing for young children in kindergarten through third grade. Others found no significant differences in vocal accuracy with or without a text with preschool children (Smale, 1988) or with kindergartners and first graders (Levinowitz, 1987).

Young children who sing accurately alone often do not sing as accurately in a group (Goetze, 1986; Smale, 1988; Rutkowski, 1990). Rutkowski (1990) carried this idea further and found that kindergarten children who have both individual and group training sing the most accurately. Studies examining positive and negative effects of harmonic accompaniment on children's vocal accuracy have produced mixed results (Petzold, 1969; Sterling, 1985; Stauffer, 1986).

Types of vocal test patterns and song materials have also been studied to determine relative difficulty of learning (Jarjisian, 1981, 1983; Sinor, 1985). Brief patterns rather than whole songs

(Flowers and Dunne-Sousa, 1989) and breath control techniques (Phillips, 1985; 1992) have also been found to contribute to greater vocal accuracy.

Several textbooks that address vocal music for young children do not discuss specific teaching strategies or research related to the teaching of songs (McDonald & Simons, 1989; Phillips, 1992; Anderson & Lawrence, 2001). Other textbooks discuss teaching strategies based on the teaching experiences of the authors without experimental justification. Campbell and Scott-Kassner (2002) suggest that the teacher should sing the song once or twice and then have the children echo either part or all of the song before the teacher joins the class in the singing. The teacher's singing then provides light support for the children's singing. Bennett (1999) concurs, stating, "A primary way to encourage children's singing is to support their voices with yours rather than to sing with boisterous or raucous volume. Once children know the song, sing with a lighter, less projecting tone, then drop out of the singing so that their voices can take responsibility to 'keep the song going'." Again, the recommendations of these authors have not been tested.

We have no research to confirm whether this "keeping the song going" encourages children to become confident singers. Do they benefit from the support of a more accurate vocal model of the teacher or do they become dependent singers and less vocally accurate without the vocal crutch of the teacher?

Feierabend (1995; 2001) disagrees with those who advocate joining the children when they sing. He refers to his "Golden Rule": "Sing <u>for</u> the class, not <u>with</u> the class." (p. 35) (original emphasis). Hackett and Lindeman (2001) soften this admonition and advise music teachers, "As soon as they are able, children should sing independently and unaccompanied, without the help of the teacher's voice or recordings." (p. 38)

These elementary music textbooks advise how music teachers should teach songs based on the experiences of the authors. None of these texts, however, cite any research to document, which is the best way for children to learn to sing songs or to learn to sing on pitch. Therefore, the purpose of our study was to conduct a controlled investigation of the effect of different methods of teaching songs to young children on vocal pitch accuracy.

Method

Upon the recommendation of the District Vocal Supervisor, three experienced music teachers were invited to participate in this study. These music educators teach in two elementary schools in south Texas with ethnically and socio-economically diverse student populations. All three elementary music educators are veteran teachers having taught between 12 and 25 years, direct children's choirs, and are interested in the development of children's vocal accuracy. They attended a training session to learn to teach songs two different ways for their classes. Teaching suggestions were prepared for the teachers to follow as a guide for the first few weeks. Parental information and permission slips were obtained from 197 children in 9 kindergarten classrooms. One hundred forty-seven kindergarten children completed both the pre- and post-test following 12 weeks of instruction.

The three teachers each had two treatment classes and one control class: *Treatment A Classes:* The music teacher only *sang for* the children and not *with* the children. Even when the song was well learned, the teacher continued to listen to the children sing and did not vocally join the class. Fifty-four children, 24 boys and 30 girls, were in Treatment A. *Treatment B Classes*: The music

teacher *sang with* the class and not *for* the class. She encouraged the class to sing along with her even while learning a new song. Treatment B consisted of 46 children, 26 boys and 20 girls. *Control Classes*: The teacher taught and reinforced songs as she deemed appropriate. Forty-seven children were in the control classes, 26 boys and 21 girls.

Children attended traditional music classes twice each week for 25 minutes in which they sang songs and participated in traditional kindergarten music activities. On Fridays, music classes for these children consisted of participating in rhythm games in the gymnasium or working on music programs in the computer lab.

During the first week of September, the Vocal Accuracy Assessment Instrument was administered. This instrument is a taped test that is administered individually to each child in a quiet room. Children are asked to echo eight three-note phrases of a complete "test song" developed by Joanne Rutkowski (1996). While the Rutkowski Singing Voice Development Measure measures children's vocal development as they mature, the Vocal Accuracy Assessment Instrument used in this study focuses on vocal pitch accuracy rather than upon vocal development. Two blinded observers analyzed recordings. Interjudge reliability was calculated at .92. Children received one point for each of 24 notes sung accurately. The Vocal Accuracy Assessment Instrument has previously been used in an earlier study with first-graders (Cousins & Persellin, 2001) and, following a pilot test, was modified for kindergartners in this study.

The voice of a talented 11 year-old singer was recorded for this assessment. Green (1987) found that children find echoing a child model to be easier than echoing an adult male, and adult female, or an instrumental model. The test model was prerecorded to assure that the same tempo (120m.m.) and accuracy of intonation was used for all children for both the pre-test and post-test.

Teachers also kept journals during the treatment period to register their comfort level and perception of the classes while following the teaching procedures. Classroom teachers also completed a questionnaire assessing the role of music in their classes.

Results

Results indicated that treatment had no significant effect on the score change (p = .2276). Vocal accuracy of the children in all three groups did not change significantly regardless of which treatment they received (Figure 1). All children in this study increased their vocal accuracy an average of 1.95 points (p = .0005). This increase was also seen with the control classes. Trends toward improvement suggested by points in Figure 1 were not statistically significant after 3 months instruction. As anticipated, children with low pretest scores increased their scores more than did children with high pretest scores (p = .0013). The lower the score on the pre-test, the greater the change was found in all three groups (Figure 2).

The nine kindergarten classroom teachers of these children also filled out a questionnaire to determine their use of singing in the classroom. They indicated that they sing with their children between 10 and 60 minutes each day with 25 minutes as the average. They reported that between 15 and 75% (average 56%) of this time was spent singing along with a cassette or CD. Three of the nine teachers described their singing voice as low; the other six classified their singing voices as medium. None of these kindergarten teachers classified their singing voice as high. They all used singing to teach or reinforce basic concepts (an average of 4.4 out of 5) and they enjoy singing with the class to help support their singing (4.3 out of 5). When asked how often they listen to the children sing rather than sing with them, they indicated an average of 2.8 (with 1 = rarely, and 5 = always).

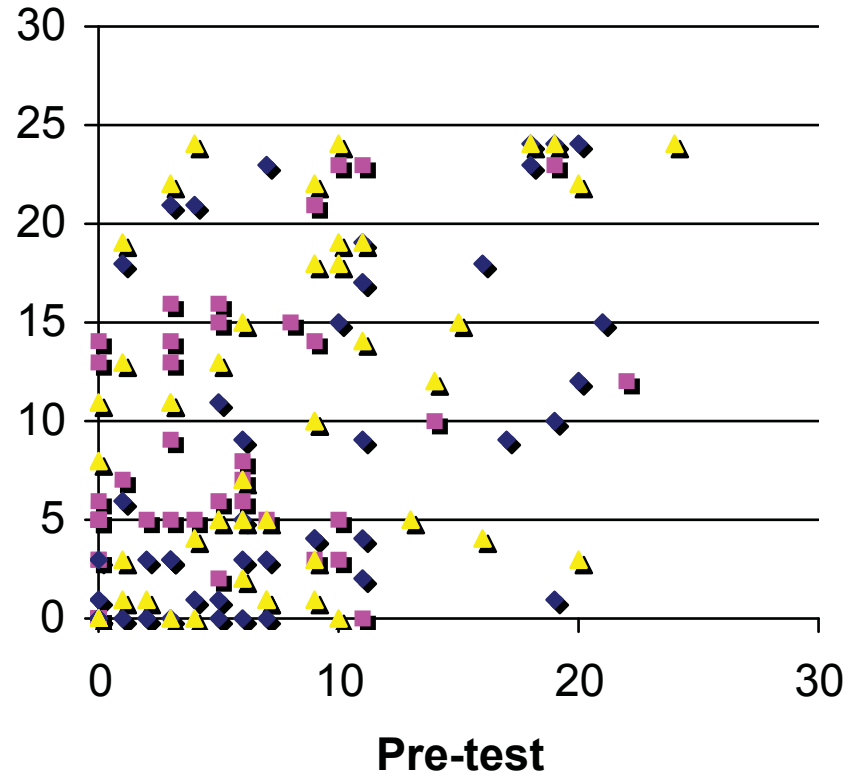

Figure 1: Comparison of Pre- and Post-tests of "Sing fc
"Sing with", and Control Groups

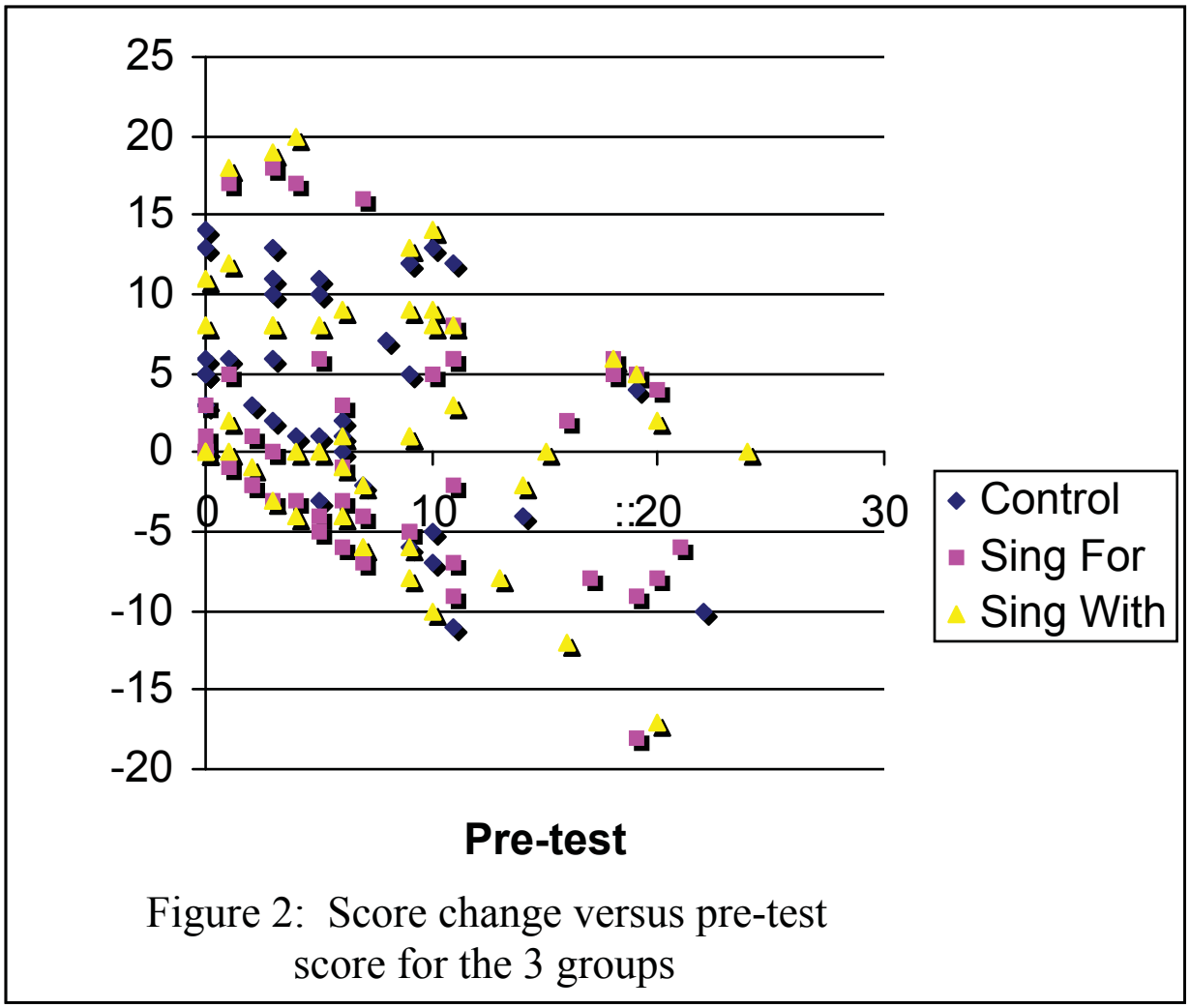

Figure 2: Score change versus pre-test score for the 3 groups

The three music teachers' self-reported assessment of their successes, challenges, and comfort levels with the teaching strategies were also analyzed for insights into teacher education strategies. As anticipated, the teachers expressed some frustration with the limits placed on them in the teaching process for this study. All three of these experienced teachers have used various combinations of singing for and with children depending on the song, the activity, the class, and what they deemed appropriate. They all found that a variety of teaching techniques including puppets helped them *sing for* or *sing with* the children. A few of the comments in their journals are included here:

"I'm feeling frustrated not being able to keep the children's focus by not echoing, just telling them to sing with me. My "split second singers" in the *sing with* class are equally good at singing with me always. I still feel some frustration because their singing is just not as accurately in-tune.

Four weeks into the study, one of the teachers wrote, "(I'm) Trying to teach fast/slow. Is the *sing with* group really getting it? Are they just following my lead? Wish I could NOT SING and find out!!! In the process of teaching new songs the *sing with* group seems to take longer to learn the tricky part of the song… Having them mirror me showing the melodic contour helps

but it is still not right. It seems like I have to do more talking with this group and it would be just so much easier to sing it for them and have them echo."

> "It takes a great deal of concentration to teach (using) only one method. It is fatiguing to sing with the entire class (with the *sing with* class) and never get a break."
>
> "I find it is difficult not to sing with the *sing for* class when we are playing a game and they can't keep the song going. It gets really quiet and they all look at me for help."
>
> "At times it feels like it is taking forever to teach a song that shouldn't take but a couple of minutes to teach (using the *sing with* method). The kids are great, though, and just rock along with me."

Teachers began to sense that perhaps one of their classes was singing more accurately than the others. While they are still expressing some frustration, they were noticing some possible effects of the treatment. Some of their comments are included here:

> "The *sing for* class overall sings very well and always listens first and matches pitch. They have really loved the echo style of teaching and listen so well."
>
> "Both groups performed well on *Apple Tree*. Pitch with *sing for* class much better than with the *sing with* group. Low 'do' is a work in progress with the *sing with* group."
>
> "The *sing for* group seems to be a little better about singing the correct notes as a class. Individuals still tend to sing it incorrectly but correct it after echoing. I also hear the classroom teacher singing with the children, but with a lower 'chesty' voice."
>
> (December): "This whole month has been spent learning Christmas songs. It was quite a challenge to only sing for the *sing for* group. I found myself wanting to sing with them. However, I believe their recall of the songs is better than the *sing with* group. They (the children in the *sing for* class) can sing on their own. I'm not sure the *sing with* group can do that since I ALWAYS SING with them. The only thing I do know is that the *sing with* group knows the words because I have them speak them. The problem is, once they start speaking them they automatically want to sing and it turns into a song so once again I am singing with them."

Conclusions

The purpose of this study was to determine the effect of two teaching models of singing songs on kindergarten vocal accuracy. In separate classes, teachers either *sang with* or *sang for* their students. All children were tested prior to the study and again after three months. Results using a vocal accuracy instrument revealed that the *sing for* group was no more accurate than the *sing with* group. Statistical analyses failed to document that one treatment was more effective than the other or was more effective than the control.

The reasons for this lack of effect may be several. It is possible the treatment duration was inadequate. Rutkowski (1996) found that 3 months was not long enough to see significant differences among treatment groups of kindergartners in her study. After 8 months of treatment,

however, significant differences were found. We are continuing our study and shall reassess after 8 months treatment.

It is also possible that young children quickly learn to adapt to whatever teaching style is presented to them when taught by skilled teachers. This phenomenon of a possible "skilled teacher effect" needs further investigation. Teachers wrote in their journals that children in the *sing for* classes were enjoying learning from that method. Children in the *sing with* may also have quickly learned to "lean" in a vocal sense on the teacher when they were singing everything together. The teachers may have felt frustrated that they couldn't stop and model the correct way to sing a vocal line, but the children enjoyed becoming "split-second singers" and learning to sing everything with or a half-second after the teacher sang each note.

Another possible reason for not demonstrating an effect of either of the two treatments could be the influence of the kindergarten classroom teachers. While these children attended music classes with trained music educators for 50 minutes each week, their classroom teachers reported singing with their students an additional 50 to 300 minutes each week. Fifty-six percent of this singing was done with pre-recorded music on the teachers' boom boxes. The effect of the music teachers' singing instruction and modeling may have had less of an impact on these children than daily singing along with a boom box or a classroom teacher singing in a low vocal range. University professors who teach music methods courses for early childhood and elementary education majors need to continue to stress the importance of singing in vocal ranges appropriate for young voices in order that children can match pitch with them. Classroom teachers also need support and guidance on criteria to use when choosing recorded music to play in their rooms for children.

In addition to various teacher style effects, extracurricular influences must be considered. Some children's singing performance in all groups was clearly superior. Many sang at home and church and had diverse music experiences other than those presented at school. Analysis of this additional data will be included in a subsequent publication following completion of 8 months' treatment.

Analysis of teachers' comments revealed frustration with the limitations on their teaching methodology as posed in this study. Their instinctive teaching styles and methods fit uncomfortably with the treatments imposed. Their teaching most likely reflected the methods used to teach them in their teacher teaching training courses and workshops as well as their years of teaching experience.

Is a specific teaching method or are several methods more effective? Whereas, some educators favor *singing for* the children (Feierabend, 1995, 2001; Hackett & Lindeman, 2001) and others recommend *singing with* the children to keep the song going (Campbell & Scott-Kassner, 2002; Bennett, 1999) neither method has been systematically studied. Perhaps the final analysis of our investigation and of other studies will reveal the most effective techniques. References

References

Anderson, W., & Lawrence, J. (2001). *Integrating music into the elementary classroom*, Belmont, CA: Wadsworth.

Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education, 32* (1), 15-24.

Bennett, P.D. (ed.) (1999). *Song play*. Milwaukee, WI: Hal Leonard Corporation.

Campbell, P.S., & Scott-Kassner, C. (2002). *Music in childhood.* (2nd ed.), Belmont, CA: Wadsworth.

Cousins, S., & Persellin, D. (1999). The effect of Curwen hand signs on vocal accuracy of young children. *Texas Music Education Research, 1999*, 17—20. Retrieved from http://www.tmea.org/080_College/Research/Cousins&Persellin.html

Feierabend, J. (1995). *First steps in music for nursery and preschool: The curriculum*. Simsbury, CT: First Steps in Music, Inc.

Feierabend, J. (September 6 - 8, 2001). *First steps in music seminar.* Judson Montessori School, San Antonio, Texas.

Flowers, P. J., & Dunne-Sousa, D. (1989). Pitch-pattern accuracy, tonality and vocal range in preschool children's singing. *Journal of Research in Music Education, 38* (2) 102-114.

Goetze, M. (1986). Factors affecting accuracy in children's singing. (Doctoral dissertation, University of Colorado, 1985). *Dissertation Abstracts International, 46,* 2955A.

Goetze, M., Cooper, N., & Brown, C. (Spring, 1990). Recent research on singing in the general music classroom. *Bulletin of the Council of Research in Music Education*, 16-37.

Gordon, E. (1979). *Primary measures of music audiation*. Chicago: G.I.A. Publications.

Green, G. (1987). The effect of vocal modeling on pitch-matching accuracy of children in grades one through six. (Doctoral dissertation, The Louisiana State University) *Dissertation Abstracts International, 48*, 1410A.

Hackett, P., & Lindeman, C. (2001). *The musical classroom*, (5th Ed.) Prentice Hall. Upper Saddle River, New Jersey.

Hendley, J. A., & Persellin, D.C. (1996). The comparative effects of the lower adult male voice and the male falsetto voice on children's vocal accuracy. *Update: Applications of Research in Music Education, 14* (2), 9 – 14.

Hermanson, L.W. (1972). An investigation of the effects of timbre on simultaneous vocal pitch acuity of young children. (Doctoral dissertation, Teachers College, Columbia University, New York, 1971) *Dissertation Abstracts International, 323*, (3558A)

Jarjisian, C. (1981). The effects of pentatonic and/or diatonic pitch pattern instruction on the rote-singing achievement of young children. *Dissertation Abstracts International, 42*, 2015A. (University Microfilms No. 8124581)

Jarjisian,C. (1983). Pitch pattern instruction and the singing achievement of young children. *Psychology of Music, 11*, 19-25.

Levinowitz, L. (1987). An experimental study of the comparative effects of singing songs with words and without words on children in kindergarten and first grade. (Doctoral dissertation, Temple University). *Dissertation Abstracts International, 48*, 863A.

McDonald, D. & Simons, G. (1989). *Musical growth and development: Birth through six.* New York: Schirmer Books.

Murphy C., & Persellin, D.C. (1993). Improving vocal accuracy of first graders through learning modalities. *Texas Music Education Research*, 56-61.

Persellin, D.C. (1993). The effect of learning modalities when used in pitch matching of patterns within complete songs. *Perceptual and Motor Skills, 76* (1), 313-314.

Persellin, D.C. (1994). Effects of learning modalities on melodic and rhythmic retention and on vocal pitch matching by preschool children. *Perceptual and Motor Skills, 78,* 1231–1234.

Petzold, R. (1969). Auditory perceptions by children. *Journal of Research in Music Education, 17* (1), 82-87.

Phillips, K. (1985). Training the child voice. *Music Educators Journal, 72* (4),19-22, 57-58.

Phillips, K. (1992). *Teaching kids to sing.* New York: Schirmer Books.

Rutkowski, J. (1996). The effectiveness of individual/small-group singing activities on kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education, 44* (4), 353-368.

Sims, W., Moore, R., & Kuhn, T.L. (1982). Effects of female and male vocal stimuli, tonal pattern length and age of vocal pitch-matching abilities of young children from England and the United States. *Psychology of Music*, Special Issue: Proceedings of the IX International Seminar of Research in Music Education, 104-108.

Sinor, E. (1985). The singing of selected tonal patterns by preschool children. (Doctoral dissertation, Indiana University, 1984) *Dissertation Abstracts International, 45*, 3299A.)

Smale, M. (1988). An investigation of pitch accuracy of four- and five-year-old singers. (Doctoral dissertation. University of Minnesota, 1987). *Dissertation Abstracts International, 48*, 2013A.)

Stauffer, S. (1986). *An investigation of the effects of melodic and harmonic context on the development of singing ability in primary grade children*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Sterling, P. (1985). A developmental study of the effects of accompanying harmonic context on children's vocal pitch accuracy of familiar melodies. *Dissertation Abstracts International, 45*, 2436A. (University microfilms No. DA8425806)

Yarbrough, C., Bowers, J., & Benson, W. (1992). The effects of vibrato on the pitch-matching accuracy of certain and uncertain singers. *Journal of Research in Music Education, 40* (1), 30–38.

Yarbrough, C., Green, G., Benson, W., & Bowers, J. (Winter, 1991). Inaccurate singers: An exploratory study of variables affecting pitch-matching. *Bulletin for the Council for Research in Music Education. 107*, 23-34.

Youngson, S.C., & Persellin, D.C. (2001). The effect of Curwen hand signs on vocal accuracy of young children. *Kodaly Envoy, 27* (2), 9-12.

Edited by
Mary Ellen Cavitt, *The University of Texas at San Antonio*

Performance Evaluation in Music

Jennifer L. Stewart
The University of Texas at Austin

Why is music performance assessed? The most frequent reason is to determine whether students have met specific criteria. Curricular objectives appear meaningless without an evaluation of students' performance of those objectives. Since the adoption of the *National Standards for Arts Education* (1994) and with the publicity associated with the *National Assessment of Educational Progress* (1997), evaluating students' performance skills has become more important than ever before.

How is performance of music literature best evaluated? What are the most important aspects of performance to be measured? What measurement instruments are reliable and valid? These and many other questions emerge when the topic of assessment is discussed. In association with the *National Standards for Arts Education* (1994), MENC published the *Performance Standards for Music* (1996). A statement in the introduction speaks to the problem of assessment in music. The statement reads, "Because there are no widely used standardized tests in music, as there are in most other basic disciplines, music educators lack a solid and uniform basis for making reliable, valid, and fair assessments of student achievement that are consistent from one setting to another and from one time to another" (p. 1).

Assessment of music performance serves many purposes and is an integral part of music education in a wide variety of settings. Assessment is also an integral part of music research. As researchers develop questions for study, they also face the challenge of developing or selecting valid, reliable instruments to measure research variables. Challenges are particularly great when researchers study questions that relate to music performance.

Systematic investigation of assessment in music research can provide insight as to the types of measurement instruments that are most commonly used and what particular elements of music are most often measured. In an article that discussed measures of instructional effectiveness, Duke (1999/2000) described the importance of the research review when he wrote, "The findings of individual investigations may remain disconnected from each other and from meaningful answers to important questions in the discipline, which is why it seems particularly important to collect and synthesize the results of research conducted over extended intervals of time in ways

that speak to educational practice and highlight consistent findings that may clarify the principles of teaching and learning" (p. 2).

The present study follows in a line of research that has used comprehensive analyses of music research to examine such areas as modes of inquiry in music therapy (Jellison, 1973), content analysis of articles in the *Journal of Research in Music Education* (Yarbrough, 1981, 1996), measurements of solo instrumental music performances (Zdzinski, 1991) analysis of the test instruments used by authors in the *Journal of Research in Music Education* (Grashel, 1996), and measures of instructional effectiveness (Duke, 1999/2000). Reviews of published research have also been conducted specific to music therapy topics and populations in Therapy and education (AMTA, 2000).

The purpose of this study was to review 30 years (1970-2000) of published, empirical literature and to identify, categorize, and count types of measurement instruments used to evaluate music performance and elements of music. This review focused on identifying the types of measurement instruments utilized by researchers and the performance elements that were measured in the performances. Secondary areas of interest included identification of the author of the measurement instrument and identification of the adjudicators and their qualifications.

Method

For this investigation, I reviewed 30 years of published, empirical music research literature (1970-2000) that evaluated the music performances of its' subjects. The subjects in the study were children and adults who sang and/or played instruments.

Music performance for this study was defined as any performance of extant literature (including jazz music), etudes, and sight-reading exercises and included performances of music literature in individual and group settings. Studies that included commercially available performance evaluations were examined as well as studies that included researcher-designed measurements.

A manual search was performed of three sources, the *Journal of Research in Music Education* (volumes 18-48), the *Bulletin of the Council for Research in Music Education* (volumes 23-145), and the *Journal of Music Therapy* (volumes 8-37). The studies resulting from the manual search were then analyzed according to the purpose(s) identified by the researcher, variables, participants, and evaluators. Types of measurement instruments used to measure performance were identified, as were specific performance elements (e.g., tone quality, rhythmic accuracy, intonation, expression, etc.). All components were subjected to frequency counts.

Results

Following the manual search, 96 studies were identified that met the criteria for the review (See Appendix). A total of 123 evaluations were found in 96 studies. Of the 123 evaluations, rating scales (n = 55) and point systems (n = 43) were used most frequently. Rating scales typically assigned a number representing the degree of agreement the adjudicator felt best described the performance or the performance element being measured. Point systems awarded or took away points based on correct or incorrect performance of the performance elements being measured. In some articles, the type of measurement instrument (n = 25) and the performance elements (n = 14) were not described in detail and could only be considered "unclear." There were 9 occurrences where performance elements were labeled as "other."

Table 1 presents frequency counts for identified performance elements. Of 471 elements measured in 123 evaluations, rhythmic accuracy was the most frequently measured (*n* = 87) followed by pitch accuracy (*n* = 46) and tempo (*n* = 39). Characteristics of sound were also measured frequently (i.e., tone quality (*n* = 32), articulation (*n* = 30), overall musical effect (*n* = 25), expression (*n* = 23), and intonation (*n* = 22)). Performance elements measured in ten studies or less were body or hand position (*n* = 10), melodic interval or contour (*n* = 9), balance and blend (*n* = 8), maintenance of key center (*n* = 8), diction (*n* = 7), other factors related to tone quality (*n* = 7), musicianship (*n* = 6), and range (*n* = 3).

Researcher-designed measurement instruments were identified in 56 studies and commercial measures were identified in 23 studies. In 41 articles, the authors did not reveal the origin of the measurement instrument. Three studies adapted previously written measurement instruments to suit their particular research needs. One of the most popular commercial measures for evaluation of student performance was the *Watkins-Farnum Performance Scale* (*n* = 15), published in 1962.

Overall there were 102 different panels of adjudicators in 96 studies and 123 evaluations. Studies with more than one evaluation often used the same panel of adjudicators. Independent judges were most often chosen to adjudicate music performance (*n* = 55), with researchers identified as adjudicators in 38 studies and both researchers and independent adjudicators identified in nine studies. Specific information regarding the adjudicators' qualifications was given in 20 articles.

Table 1
Types of Performance Elements Measured and Types of Evaluations Found in 96 Research Studies.

Performance Elements	Frequency
Rhythmic Accuracy	87
Pitch Accuracy	46
Tempo	39
Tone Quality	32
Articulation	30
Overall Musical Effect	25
Expression	23
Intonation	22
Hesitations, Repeated Notes, Note Omission	19
Phrasing	16
Dynamics	14
Repeats	14

Unclear	14
Technique	13
Style Interpretation	10
Melodic Interval or Contour	9
Other	9
Balance and Blend	8
Key Center or Tonal Center	8
Diction	7
Other Factors Related to Tone Quality	7
Musicianship	6
Range	3
Total	471

Table 1 *Continues*

Table 1 *continued*

Types of Evaluation Instruments	Frequency
Rating Scales	55
Point Systems	43
Unclear	25
Total	**123**

Discussion

The first guideline for assessment stated in *Performance Standards for Music* (MENC, 1996) is, "Assessment should be standards-based and should reflect the music skills and knowledge that are most important for students to learn" (p. 7). In teaching, assessment of performance skills is an important component in measuring whether specific curricular objectives have been met. In research, decisions regarding assessment are also tied to objectives—research objectives and research questions.

Research questions were not examined in this review and it was assumed that the measurement instruments used in the 96 studies were appropriate to the questions asked. It may be that the research questions were more effectively answered by measuring the performances of individuals since only 10 of 123 evaluations in 96 studies measured group performances. The frequency of individual measurement also suggests that there are few recognized instruments for group adjudication.

The 123 assessments in the 96 research studies varied greatly. Most of the variety was among researcher-designed instruments. Researcher-designed instruments comprised 45.52% of the measurement instruments in this study and commercial measures of music performance, such as the *Watkins Farnum Performance Scale* or the adjudication form by the *Music Educators National Conference,* comprised 18.69% of the measurement instruments. The large number of measurement instruments designed by researchers may be due to the variety of research questions asked: researchers designed their own measures to study their questions of interest. Across the studies, there are few examples of standardized music performance assessments that measure the same performance elements.

By far, the most frequently measured performance element in the research literature was rhythmic accuracy (n = 87). Rhythmic accuracy may have been measured more frequently than other performance elements because rhythm may be more easily quantified than other performance elements. Another reason for measuring rhythmic accuracy frequently may be that researchers asked questions that can best be answered by measuring that performance element. Perhaps broadening the scope of the research questions may provide researchers more frequent

opportunities to measure other performance elements such as tone quality, intonation, or style characteristics.

In summary, based on an analysis and synthesis of the content of 96 articles, researchers most frequently use rating scales and point systems to evaluate performance. The measurement instruments were most often researcher-designed and used to measure a variety of performance elements, although rhythmic accuracy was measured most frequently. We can assume that researchers employed the most appropriate measurement instruments available and measured the performance elements that were directly related to their research questions.

Despite the fact that the music performance of individuals has been widely assessed throughout the last 30 years of published research, only a few performance elements are consistently included in the evaluations and there is little consistency in the use of measurement instruments. Excellence in music performance is a goal of music education, although it appears that there is still much work to be done if valid, reliable, generic measures of student performance are to be developed. Also to be examined are ways in which the accuracy, quality, and character of group instrumental and vocal performances can best be measured.

References

Duke, R. A. (1999/2000). Measures of Instructional Effectiveness in Music Research. *Council* for Research in Music Education, 143, 1-48.

Grashel, J. (1996). Test instruments used by *Journal of Research in Music Education* authors from 1980-1989. *Update, 15,* 24-30.

Jellison, J. A. (1973). The frequency and general mode of inquiry of research in music therapy, 1952-72. *Council for Research in Music Education, 35,* 1-8.

National Center for Education Statistics. (2001). *National Assessment of Educational Progress.* Retrieved 8/16/2001 from http://www.nces.ed.gov/nationsreportcard/.

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

Performance standards for music: Strategies for assessing progress toward the National Standards, grades preK-12. (1996). Reston, VA: Music Educators National Conference.

The American Music Therapy Association. (2000). *Effectiveness of Music Therapy Procedures: Documentation of Research and Clinical Practice* (3rd ed.). Silver Spring, MD: author.

Yarbrough, C. (1984). A content analysis of the Journal of Research in Music Education, 1953-1983. *Journal of Research in Music Education, 32,* 233-242.

Yarbrough, C. (1996). The future of scholarly inquiry in music education: 1996 Senior Researcher Award Acceptance Address. *Journal of Research in Music Education, 44,* 190-203.

Zdzinski, S. F. (1991). Measurement of solo instrumental music performance: a review of literature. *Council for Research in Music Education 109,* 47-58.

Articles Reviewed

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

Alexander, L. & Dorow, L. G. (1983). Peer tutoring effects on the music performance of tutors and tutees in beginning band classes. *Journal of Research in Music Education, 31,* 33-47.

Alford, D. L. (1971). Emergence and development of music responses in preschool twins and singletons: a comparative study. *Journal of Research in Music Education, 19,* 222-226.

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

Apfelstadt, H. A. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education, 32,* 15-24.

Appel, S. S. (1976). Modifying solo performance anxiety in adult pianists. *Journal of Music Therapy, 13,* 2-16.

Austin, J. R. (1988). The effect of music contest format on self-concept, motivation, achievement, and attitude of elementary band students. *Journal of Research in Music Education, 36,* 95-107.

Azzara, C. D. (1993). Audiation-based improvisation techniques and elementary instrumental students' music achievement. *Journal of Research in Music Education, 41,* 328-342.

Bergee, M. J. (1993). A comparison of faculty, peer and self-evaluation of applied brass jury performances. *Journal of Research in Music Education, 41,* 19-27.

Bergee, M. J. (1995). Primary and higher-order factors in a scale assessing concert band performance. *Council for Research in Music Education, 126,* 1-14.

Bergee, M. J. (1997). Relationships among faculty, peer and self-evaluations of applied performances. *Journal of Research in Music Education, 45,* 601-612.

Bergonzi, L. (1997). Effects of finger markers and harmonic context on performance of beginning string students. *Journal of Research in Music Education, 45,* 197-211.

Betts, S. L. and Cassidy, J. W. (2000). Development of harmonization and sight-reading skills among university class piano students. *Journal of Research in Music Education, 48,* 151-161.

Boyle, J. D. (1970). The effect of prescribed rhythmical movements on the ability to read music at sight. *Journal of Research in Music Education, 18,* 307-318.

Caimi, F. J. (1981). Relationships between motivation variables and selected criterion measures of high school band directing success. *Journal of Research in Music Education, 29,* 183-198.

Cassidy, J. W. (1993). Effects of various sightsinging strategies on nonmusic majors' pitch accuracy. *Journal of Research in Music Education, 41,* 293-302.

Chinn, B. J. (1997). Vocal self-identification, singing style and singing range in relationship to a measure of cultural mistrust in African-American adolescent females. *Journal of Research in Music Education, 45,* 639-649.

Coffman, D. D. (1990). Effects of mental practice, physical practice, and knowledge of results on piano performance. *Journal of Research in Music Education, 38,* 187-196.

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

Davis, A. P. (1998). Performance achievement and analysis of teaching during choral rehearsals. *Journal of Research in Music Education, 46,* 496-509.

Delzell, J. K. (1989). The effects of musical discrimination training in beginning instrumental music classes. *Journal of Research in Music Education, 37,* 21-31.

Demorest, S. M. (1998). Improving sight-singing performance in the choral ensemble: the effect of individual testing. *Journal of Research in Music Education, 46,* 182-192.

Demorest, S. & May, W. V. (1995). Sight-singing instruction in the choral ensemble: factors related to individual performance. *Journal of Research in Music Education, 43,* 156-167.

Dileo, C. L. (1976). The relationship of diagnostic and social factors to the singing ranges of institutionalized mentally retarded persons. *Journal of Music Therapy, 13,* 17-28.

Dorow, L. G., & Greer, R. D. (1977). The reinforcement value of a music instrument for beginning instrumentalists and the influence of discovery versus teacher approval on achievement. *Journal of Music Therapy, 14,* 2-16.

Duerkson, G. L. (1972). Some effects of expectation on evaluation of recorded musical performance. *Journal of Research in Music Education, 20,* 268-272.

Dunn, D. E. (1997). Effect of rehearsal hierarchy and reinforcement on attention, achievement, and attitude of selected choirs. *Journal of Research in Music Education, 45,* 547-567.

Eckholm, E. (2000). The effect of singing mode and seating arrangement on blend and overall choral sound. *Journal of Research in Music Education, 48,* 123-135.

Elliott, C. A. (1995). Race and gender as factors in judgments of musical performance. *Council for Research in Music Education, 127,* 50-55.

Fiske, Jr., H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education, 23,* 186-196.

Fiske, Jr., H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education, 23,* 256-263.

Flom, J. H. (1971). An investigation of growth in musical facts and concepts, musical discrimination, and vocal performance proficiency as a result of senior high school music experiences. *Journal of Research in Music Education, 19,* 433-442.

Flowers, P. J. & Dunne-Sousa, D. J. (1990). Pitch-pattern accuracy, tonality, and vocal range in preschool children's singing. *Journal of Research in Music Education, 38,* 102-114.

Froehlich, H. (1977). The relationship of selected observational variables to the teaching of singing. *Journal of Research in Music Education, 25,* 115-130.

Froehlich, H. (1979). Replication of a study on teaching singing in the elementary general music classroom. *Journal of Research in Music Education, 27,* 35-45.

Froseth, J. O. (1971). Using MAP scores in the instruction of beginning students in instrumental music. *Journal of Research in Music Education, 19,* 98-105.

Garofalo, R. J. & Whaley, G. (1979). Comparison of the unit study and traditional approaches for teaching music through school band performance. *Journal of Research in Music Education, 27,* 137-142.

Geringer, J. M., and Madsen, C. K. (1998). Musicians' ratings of good versus bad vocal and string performances. *Journal of Research in Music Education, 46,* 522-534.

Goolsby, T. W. (1999). A comparison of expert and novice music teachers' preparing identical band compositions: an operational replication. *Journal of Research in Music Education, 47,* 174-187.

Gordon, E. (1970). First year results of a five-year longitudinal study of the musical achievement of culturally disadvantaged students. *Journal of Research in Music Education, 18,* 195-21.

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education, 42,* 105-114.

Grutzmacher, P. A. (1987). The effect of tonal pattern training on the aural perception, reading recognition, and melodic sight-reading achievement of first-year instrumental music students. *Journal of Research in Music Education, 35,* 171-181.

Hamann, D. L. (1982). An assessment of anxiety in instrumental and vocal performances. *Journal of Research in Music Education, 30,* 77-90.

Kelly, S. N. (1997). Effects of conducting instruction on the musical performance of beginning band students. *Journal of Research in Music Education, 45,* 295-305.

Kendall, M. J. (1988). Two instructional approaches to the development of aural and instrumental performance skills. *Journal of Research in Music Education, 36,* 205-219.

Killian, J. N. (1991). The relationship between sight-singing accuracy and error detection in junior high singers. *Journal of Research in Music Education, 39,* 216-224.

Killian, J. N. (1996). Definitions of "knowing": comparison of verbal report versus performance of children's songs. *Journal of Research in Music Education, 44,* 215-228.

Klinedinst, R. E. (1991). Predicting performance achievement and retention of fifth grade instrumental students. *Journal of Research in Music Education, 39,* 225-238.

Kostka, M. J. (2000). The effects of error-detection practice on keyboard sight-reading achievement of undergraduate music majors. *Journal of Research in Music Education, 48,* 114-122.

LeBlanc, A., Jin, Y. C., Obert, M., Siivola, C. (1997). Effect of audience on music performance anxiety. *Journal of Research in Music Education, 45,* 480-496.

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

MacKnight, C. B. (1975). Music reading ability of beginning wind instrumentalists after melodic instruction. *Journal of Research in Music Education, 23,* 23-34.

Madsen, C. K. & Geringer, J. M. (1981). The effect of a distraction index on improving practice attentiveness and musical performance. *Council for Research in Music Education, 66-67,* 46-52.

Madura, P. D. (1996). Relationships among vocal jazz improvisation achievement, jazz theory knowledge, imitative ability, musical experience, creativity, and gender

Marchand, D. J. (1975). A study of two approaches to developing expressive performance. *Journal of Research in Music Education, 23,* 14-22.

McCarthy, J. F. (1974). The effect of individualized instruction on the performance achievement of beginning instrumentalists. *Council for Research in Music Education, 38,* 1-16.

McCarthy, J. F. (1980). Individualized instruction, student achievement, and dropout in an urban elementary instrumental music program. *Journal of Research in Music Education, 28,* 59-69.

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

McPherson, G. (1993-94). Evaluating improvisational ability of high school instrumentalists. *Council for Research in Music Education, 119,* 11-20.

McPherson, G. (1994). Factors and abilities influencing sightreading skill in music. *Journal of Research in Music Education, 42,* 217-231.

McPherson, G. E. (1995-96). Five aspects of musical performance and their correlates. *Council for Research in Music Education, 127,* 115-121.

McPherson, G. E. (1997). Cognitive strategies and skill acquisition in musical performance. *Council for Research in Music Education, 133,* 64-71.

McPherson, G. E., Bailey, M., Sinclair, K. E. (1997). Path analysis of a theoretical model to describe the relationship among five types of musical performance. *Journal of Research in Music Education, 45,* 103-129.

Mills, J. (1987). Assessment of solo musical performance- a preliminary study. *Council for Research in Music Education, 91,* 119-125.

Mizener, C. P. (1993). Attitudes of children toward singing and choir participation and assessed singing skill. *Journal of Research in Music Education, 41,* 233-245.

Mota, G. (1997). Detecting young children's musical aptitude: a comparison between standardized measures of music aptitude and ecologically valid musical performances. *Council for Research in Music Education, 133,* 89-94.

Noble, R. F. (1971). Effects of a concept teaching curriculum on performance achievement in elementary school beginning bands. *Journal of Research in Music Education, 19,* 209-215.

Orman, E. K. (1998). Effect of interactive multimedia computing on young saxophonists' achievement and attitude. *Journal of Research in Music Education, 46,* 62-74.

Pembrook, R. G., & Robinson, C. R. (1997). The effect of mode of instruction and instrument authenticity on children's attitudes, information recall and performance skill for music from Ghana. *Council for Research in Music Education, 133,* 115-120.

Petters, R. (1976). Student participation in decision-making processes concerning musical performance. *Journal of Research in Music Education, 24,* 177-186.

Pierce, M. A. (1992). The effects of learning procedure, tempo and performance condition on transfer of rhythm skills in instrumental music. *Journal of Research in Music Education, 40,* 295-315.

Price, H. E. (1983). The effect of conductor academic task presentation, conductor reinforcement, and ensemble practice on performers' musical achievement, attentiveness, and attitude. *Journal of Research in Music Education, 31,* 245-257.

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.

Ramsey, J. H. (1983). The effects of age, singing ability, and instrumental experiences on preschool children's melodic perception. *Journal of Research in Music Education, 31,* 133-145.

Roberts, E., & Davies, A. D. M. (1975). Poor pitch singing: response of monotone singers to a program of remedial training. *Journal of Research in Music Education, 23,* 227-239.

Rohwer, D. (1998). Effect of movement instruction on steady beat perception, synchronization, and performance. *Journal of Research in Music Education, 46,* 414-424.

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,* 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,* 250-257.

Ross, S. L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. *Journal of Research in Music Education, 33,* 221-230.

Sang, R. C. (1987). A study of the relationship between instrumental music teachers' modeling skills and pupil performance behaviors. *Council for Research in Music Education, 91,* 155-159.

Saunders, T. C., & Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *Journal of Research in Music Education, 45,* 259-272.

Sehmann, K. H. (2000). The effects of breath management instruction on the performance of elementary brass players. *Journal of Research in Music Education, 48,* 136-150.

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

Sims, W. L. (1995). Children's ability to demonstrate music concept discriminations in listening and singing. *Journal of Research in Music Education, 43,* 204-221.

Ten Eyck, S. G. (1985). The effect of programmed materials on the vocal development of selected children's choruses. *Journal of Research in Music Education, 33,* 231-246.

Wapnick, J., Darrow, A. A., Kovacs, J., and Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education, 45,* 470-479.

Wapnick, J., Flowers, P., Alegant, M., & Jasinskas, L. (1993). Consistency in piano performance evaluation. *Journal of Research in Music Education, 41,* 282-292.

Wapnick, J., Mazza, J. K., Darrow, A. A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education, 46,* 510-521.

Wapnick, J., Mazza, J. K., Darrow, A. A. (2000). Effects of performer attractiveness, stage behavior, and dress on evaluation of children's piano performance. *Journal of Research in Music Education, 48,* 323-336.

Watkins, J. G. & Farnum, S. E. (1962). *Watkins-Farnum performance scale: Forms A and B.* Winona, MN: Leonard Music.

Webster, P. R. (1979). Relationship between creative behavior in music and selected variables as measured in high school students. *Journal of Research in Music Education, 27,* 227-242.

Welch, G. F., Sergeant, D. C., & White, P. J. (1995-96). The singing competencies of five-year-old developing singers. *Council for Research in Music Education, 127,* 155-162.

Welch, G. F., Sergeant, D. C., & White, P. J. (1997). Age, sex, and vocal task as factors in singing "in tune" during the first years of schooling. *Council for Research in Music Education, 133,* 153-160.

Whitener, W. T. (1982). Comparison of two approaches to teaching beginning band. *Journal of Research in Music Education, 30,* 229-235.

Wolpow, R. I. (1976). The independent effects of contingent social and academic approval upon the musical on-task and performance behaviors of profoundly retarded adults. *Journal of Music Therapy, 13,* 29-38.

Young, W. T. (1971). The role of musical aptitude, intelligence, and academic achievement in predicting the musical attainment of elementary instrumental music students. *Journal of Research in Music Education, 17,* 385-398.

Zdzinski, S. F. (1992). Relationships among parental involvement, music aptitude and musical achievement of instrumental music students. *Journal of Research in Music Education, 40,* 114-125.

Zdzinski, S. F. (1996). Parental involvement, selected student attributes, and learning outcomes in instrumental music. *Journal of Research in Music Education, 44,* 34-48.