THE EFFECTS OF SELF-ASSESSMENT AND MODEL-LISTENING ON MIDDLE SCHOOL CHORUS STUDENTS' ACHIEVEMENT

by

Jenny Jane Ward

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Music

Spring 2010

THE EFFECTS OF SELF-ASSESSMENT AND MODEL-LISTENING

ON MIDDLE SCHOOL CHORUS STUDENTS' ACHIEVEMENT

by

Jenny Jane Ward

Approved: _____

Krystal Rickard, M.M.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Paul D. Head, D.M.A.
Chairperson of the Department of Music

Approved: _____

George Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

Debra Hess Norris, M.S.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I want to thank my thesis advisor, Professor Krystal Rickard, for all of her help and guidance with my study and research. I would also like to thank Dr. Suzanne Burton for her assistance with the drafting of my proposal. I have been amazed at the amount of time both professors have been willing to spend working with me on this.

Thank you to my thesis committee, Professor Krystal Rickard, Dr. Suzanne Burton, and Dr. Melanie DeMent, for dedicating their time and effort, and even more so for all of their words of encouragement.

I greatly appreciate the help Dr. Thomas Ilvento has provided with data analysis.

I wish to thank Sarah Aherne and Angela Sheik for their participation in this study and for all of the time they have devoted.

Finally, I wish to express my deepest gratitude to my husband, Aaron. In addition to a considerable amount of technological assistance, Aaron has been incredibly supportive and encouraging throughout the entire process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of this ten-week study was to determine the effects of self-assessment and model-listening on middle school chorus students' performance achievement. Sixty-four middle school chorus students were assigned to one of the following four groups: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment. Once per week, self-assessment groups completed self-assessments and model-listening groups listened to a model recording. During week ten, each group made final recordings which were evaluated by two judges. Data were analyzed using descriptive statistics to determine differences among the groups for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement. Data indicated that middle school choral students benefit most from no-treatment in the performance sub-areas of Tone Quality, Intonation, and Expression, and in Overall Performance. Also, self-assessment alone is the most effective treatment for improving Balance and Blend.

CHAPTER 1

REVIEW OF LITERATURE

The topic of assessment has been an important element in the education field over the past several years. Educators are not only gauging students' achievement at the end of an instructional unit or course, but also incrementally so that instructional adjustments may be made. Students are more likely to master objectives if instructional methods have been tailored to fit their learning needs. Moreover, many educators are empowering students to be a part of this formative assessment process. As defined by McMillan and Hearn (2009), "Self-assessment occurs when students judge their own work to improve performance as they identify discrepancies between current and desired performance. Self-assessment also identifies further learning targets and instructional strategies (correctives) students can apply to improve achievement." (p. 39) Student self-assessment has proven not only to increase the likelihood of success in the classroom, but also to motivate student learning. While self-assessment has customarily been used in academic or *core* classes, its use is now more widespread in arts education, particularly in the music classroom.

Assessment in the arts has received attention from the well-known Harvard research group "Project Zero" founded in 1967. In 1985, Project Zero collaborated with the Education Testing Service and the Pittsburgh Public Schools to form the large-scale project Arts PROPEL. The goal of Arts PROPEL was to create assessment tools for use

in late elementary through high school music, visual art, and imaginative writing classes with a focus on production, perception, and reflection. A list of competences was established for each of the three arts areas and then a set of curriculum-compatible exercises was created for each competency referred to as a *domain project*. These resulting domain projects came to be used within actual school curriculums and programs (Gardner, 1989).

One of the music domain projects involved the use of portfolio assessment in music. These portfolios were designed to house students' periodic self-assessments and/or reflections of ensemble performances. Students were asked to listen and reflect upon previously taped performances in an effort to gauge their progress. As a result, students' critical listening skills were greatly improved and they were able to determine the strategies necessary to improve individual and group performance achievement for themselves (Gooslby, 1995).

Research suggests that employing self-assessment in the ensemble curriculum yields many positive outcomes. Gooslby (1999) explained that self-assessment, whether done individually or as a group, is a crucial part of student independent musicianship—a major goal in music education. Burrack (2002) believed the process of becoming proficient in self-assessment leads students to be more meticulous and thoughtful musicians. According to Stamer (2002), students should not be merely told what to do or fix in the music rehearsal, they should be guided toward the solution and the responsibility of musical decision-making should be theirs.

Student self-assessment in the music classroom is a rare topic in the music education research field. The number of documented studies is small and those studies which can be found pertain specifically to instrumental ensembles. We have seen the positive effects student self-assessment can have on instrumental ensemble. The music education profession could use more information on the effects of this teaching tool in the choral spectrum.

Research purpose and questions

With the intent on improving choral music instruction, the purpose of this study was to determine the effects of routine self-assessment and model-listening on choral achievement. Research questions concerning assessment in a middle school choral ensemble are as follows:

1. Does routine student self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

2. Does routine model-listening improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

3. Does the combination of routine model-listening and self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

The following will cover terms and definitions as well as research pertaining to (a) self-assessment in the choral setting, (b) self-assessment in the instrumental setting, and (c) model-listening.

Terms and definitions

Three terms necessary to define are measurement, assessment, and evaluation. These terms will be seen periodically in the following literature overview and description of the present study. Even though these terms are related, their meanings are different. As defined by Kizlik (2009), measurement is "the process by which the attributes or dimensions of some physical object are determined" (p.1). Measurements may also be taken of a person's IQ, attitude, or preferences. Assessment is process of comparing a measurement to a known objective or goal. For example, teachers compare measurements of student knowledge to their established objectives, often in the form of a test. The most complex of the three, evaluation, involves obtaining information and using it to make a judgment. For example, a teacher would be evaluating if he or she used test score information to inform a decision on future instructional plans. Participants in the present study collected information about their singing relative to a goal (choral performance achievement) although they did not use this information to adapt future rehearsal strategies. Therefore, this variable was referred to as *self-assessment*.

Choral self-assessment

Since the idea of student self-assessment is a relatively new concept, it is intriguing that an article about self-assessment of high school chorus students was published as early as 1948. Molnar (1948) wrote the article "Self-assessment by Students" which explored the idea of having students assess their own choral technique, voice production, music reading skills, and attitude on a regular basis. After having used

this system for several years, Molnar concluded it improves student attitude toward self-improvement and more firmly fixes aims and objectives.

Robinson (1990) published an article comparing choral assessment abilities of students with varying amounts of musical experience. Surprisingly, trained and untrained musicians alike have an aural sensitivity to many of the same performance sub-areas such as overall choral sound and diction. There are other aspects that lesser trained musicians may pay little attention to, such as Intonation and balance. However, training has the tendency to significantly improve critical listening skills. Robinson suggests that small groups of students listen to ensemble performance and assess the choral ensemble on a routine basis. Over time, the whole ensemble will become more deeply engaged in listening and assessment.

Matheny (1994) proposed that student self-assessment in junior high and high school performing groups provides a solid basis for student improvement and student/director communication and understanding. Matheny provided choral students with a self-assessment form which measured attendance, contribution to the ensemble, effort, attitude, willingness to adapt one's schedule to group needs, musical skill, technical skill, performance skill, and desire for excellence. Students were asked to rate each of these areas on a scale from one to ten and then their ratings were compared with those of the teacher. Matheny concluded that the assessment process serves a valuable function by helping the teacher and students talk about and assess their performance in the group.

Darrow and Marsh (2006) conducted a study to determine if choral students were able to predict and assess their sight-singing skills. For the study, 50 community children's chorus members completed a self-rating sheet before and after sight-singing five musical examples. The authors found that students were able to predict their sight-singing proficiency with a reasonable amount of accuracy. Their performance evaluations were typically more accurate than their predictions, particularly for those students with more musical experience.

A similar article was published the same year by Barefoot (2006) in which he examined the ability of developing singers to analyze their own singing in such terms as vowel sound and focused tone. Barefoot explains that with guidance, singers can have great success in self-assessment. More importantly, singers who utilize self-assessment methods will make better use of their practice time and will develop their vocal skills more quickly. Traditional teacher-directed methods, in addition to being more time consuming, are often quite ineffective in that instructions are often subjective and can be limiting to singers with different vocal perceptions. Barefoot explained that students will grow as musicians through the use of self-assessment and this process will yield even more positive results with the incorporation of aural models.

McCall (2007) conducted a study on portfolio assessment in middle school chorus to determine how portfolios inform student and teacher learning. Involved in the study were 53 eighth grade students from a small suburban school in New York. During the course of this six-month study, students created portfolios containing questionnaires as well as individual and group video recordings. Through assessing portfolio

documentation and conducting interviews, McCall found that students' self-assessment confidence increased, the recordings were an effective tool for allowing mutual student-to-peer and student-to-teacher assessment, and students became better able to express their musical preferences.

The first important point made in the above summary of choral research is that choral students can accurately self-assess and, with training, their ability to self-assess can improve. Incorporation of aural models can increase self-assessment accuracy as well. Some of the benefits of self-assessment in the choral ensemble are that (a) students are more motivated, (b) they have a better understanding of the objectives, (c) student-director communication is improved, (d) practice time is better utilized, and, perhaps most importantly, (e) independent musicianship is promoted.

Studies show that chorus students have the capability to accurately self-assess their own choral achievement and that accurate self-assessment has many benefits such as increased motivation and objective mastery. There is little documentation, however, showing that accurate self-assessment can yield a higher level of ensemble choral achievement.

Instrumental self-assessment

The following overview will cover self-assessment research in the instrumental ensemble since this where the majority of self-assessment research has been conducted. Goolsby (1999) published an article on musical assessment in which he explained that band and orchestra teachers often waste rehearsal time by repeatedly reminding students of the same problems. One way to alleviate this problem, according to Goolsby, is to

involve students in the assessment process and approach the goal of musical independence. Goolsby suggests that ensemble students keep a copy of a self-assessment form in their folders for use on a weekly basis. Through the regular use of student self-assessment, rehearsal time will be conserved and students' performance skills will be improved.

Hewitt (2001) investigated the effects of modeling, self-assessment, and self-listening on junior high instrumentalists' music performance and attitude toward practice. Performance sub-areas included tone, Intonation, technique/articulation, melodic accuracy, Rhythmic Accuracy, tempo, and interpretation. For the study, 82 woodwind, brass, and percussion students were randomly assigned to eight treatment groups containing combinations of two modeling conditions, two self-assessment conditions, and two self-listening conditions. This pretest/posttest study was implemented over a nine-week period. Pre-performances and post-performances were compared and practice attitude questionnaires were analyzed. Relationships were determined among two modeling conditions, two self-listening conditions and two self-assessment conditions. Hewitt found no statistical significance for the four-way interaction of model, self-listening, self-assessment, and test scores. There were, however, two statistically significant three-way findings. These two groupings were: modeling, self-listening and self-assessment, and modeling, self-assessment, and test. Overall, results showed that self-assessment combined with model-listening yielded improved performance achievement over those who did not listen to a model in the areas of tone, melodic accuracy, Rhythmic Accuracy, interpretation, and Overall Performance. Intonation,

technique/articulation, and tempo, however, were aspects that did not improve. Model-listening only groups improved their performance more than students who did not listen to a model in the areas of tone, technique/articulation, Rhythmic Accuracy, tempo, interpretation, and Overall Performance but not Intonation or melodic accuracy.

Hewitt (2002) conducted a similar study one year later on junior high instrumentalists to (a) determine the nature of self-assessment tendencies, (b) examine the process of self-assessment with and without the use of a model and (c) determine whether there is a correlation between self-assessment accuracy and music performance achievement. Forty-one woodwind, brass, and percussion students (14 low-ability, 11 middle-ability, and 16 high-ability) participated in this six-week, pretest/posttest study. Students were assigned to one of two treatments groups—one with and one without an aural model. The dependent variables were performance achievement and self-assessment accuracy. At the end of the study, Hewitt compared pre-performances and post-performances and analyzed student self-assessments. Results showed that the students' self-assessment scores did not improve over time and model-listening did not aid in self-assessment accuracy. In fact, model-listening actually decreased Intonation accuracy. Finally, the students' self-assessment scores had no significant correlation with performance achievement.

Kruse (2006) was also interested in student self-assessment tendencies. He researched the effect of providing assessment rubric instruction on the accuracy of self-assessment. In this study, 36 sixth grade band students performed and recorded a piece and then self-assessed their Rhythmic Accuracy. The band was then divided into two

groups, one receiving assessment rubric instruction and the other not. The group that received instruction on using the self-assessment rubric was more accurate in their self-assessment. Kruse described self-assessment rubric instruction as a "practice strategy toward students' independent musicianship" (p. 2).

There are some similarities between the self-assessment findings in choral and instrumental settings: rehearsal time is conserved, performance achievement is improved, and musical understanding and motivation are increased. One finding that was not corroborated in the choral realm of research on student self-assessment was the correlation between self-assessment and performance achievement as seen in Hewitt's (2001) study. Other studies, including Hewitt's (2002) second study, demonstrated no correlation between the self-assessment and performance achievement.

Another finding that differed from the choral to the instrumental studies was that student self-assessment accuracy may not significantly improve over time. On the other hand, Kruse (2006) found that students' self-assessment accuracy *could* be improved with instruction. Also notable in the instrumental self-assessment literature is that self-assessment combined with model-listening yields improved performance achievement.

Model-listening

Some researchers have found model-listening to be a useful instructional strategy in the music classroom. All studies specific to model-listening were conducted in the instrumental setting. Anderson (1981) conducted research on the effects of tape-recorded aural models on sight-reading and performance skills of student clarinet players. Anderson analyzed pitch reading, rhythm reading, tempo accuracy, and Intonation

accuracy. The results, which were quite different from Anderson's predictions, showed no significant correlation between model-listening and sight-reading or performance skills.

A few years later, a study was conducted on the relative effects of guided model, guide only, and practice only treatments on accuracy of advanced instrumentalists' musical performance by Rosenthal (1984). Performance accuracy was measured in terms of correct notes, rhythm, dynamics, tempo, and phrasing/articulation. The study conditions, assigned randomly to 44 college music education students, were (a) aural/verbal model, (b) aural model, (c) verbal instructions, and (d) practice only. The highest scoring of the four groups was the model-only groups. The aural-model students scored higher than those in the guide-only and practice-only groups.

Rosenthal, Wilson, Evans, & Greenwalt (1988) researched advanced instrumentalists to examine the effects of different practice conditions on performance accuracy. Performances were assessed the same way as in Rosenthal's 1984 study. The five practice sessions used were modeling, singing, silent analysis, free practice, and control (no treatment). Sixty college music students were assigned to one of the five conditions and asked to perform a piece of music after a brief practice session. Rosenthal et al. found that listening to a model is as beneficial as actual practice for the college students. Silent analysis did not seem to be any more beneficial than sight-reading with the exception of rhythmic performance.

Morrison, Montemayor, & Wiltshire (2004) were interested in the effect of a recorded model on junior high and high school band students' performance self-

assessments, achievement, and attitude. During this five-week study, directors of three middle/junior high and two high school bands used professional recordings as part of their musical instruction for five selected pieces. Each week students listened to recordings and they completed progress reports in which they evaluated notes/rhythms, articulation/dynamics, tuning, and balance. Pre-treatment and post-treatment recordings were compared to assess ensemble achievement. Results showed no differences in ensemble achievement between model and no-model conditions. There was, however, a correlation between model and self-assessment. Students demonstrated a higher level of accuracy in their assessment of their ensemble's improvement over time with a model. They also offered a greater number of comments in the model situation. High school students tended to be more positive in their evaluations of their own performance as opposed to that of the whole ensemble and their evaluations over time showed more change. The younger students in the study had a more positive attitude toward the model pieces. All students provided more free-response comments for model pieces.

In summarizing the research conducted on model-listening, with the exception of Rosenthal's studies, there was no direct correlation between model-listening and performance achievement. Both of Rosenthal's studies did suggest that performance achievement increased with model-listening. However, these studies involved college students who already had performance experience. The most important finding here, similar to the studies on choral and instrumental self-assessment, is that model-listening can improve student self-assessment accuracy in that students are more conservative and

specific in their analyses. This information further emphasizes the need to include the aspect of model-listening in the present study.

CHAPTER 2

REVIEW OF RELATED LITERATURE

The focus of this study was to investigate the effects of routine self-assessment and model-listening on middle school chorus students' achievement. The following overview will summarize literature closely related to the present study in the areas of choral self-assessment, instrumental self-assessment, and model listening in the instrumental setting.

Choral Self-assessment

McCall (2007) conducted a study to see how portfolio assessment could inform student and teacher learning in a middle school chorus setting. The 53 eighth grade students participating in this study were members of a non-audition, elective choir at a small, suburban school. Most of these students had some prior choral experience.

Data collection. The researcher collected data for this study, primarily portfolio artifacts, from the beginning of the school year to the middle of the third marking period. Two main sets of artifacts were collected during the study, the first of which the researcher referred to as the "baseline" data. The purpose of this first data set was to help the researcher become familiar with students' goals and prior knowledge and also to help the students get acquainted with the assessment process. This data set consisted of two questionnaires, an individual performance video, and an ensemble performance video. The two questionnaires were a school and community survey, and a musicianship goal

questionnaire. For the individual performance videos, students learned and performed a simple song for the purposes of establishing videotaping protocol and designating each student's range and tessitura. Students were asked to complete self and peer evaluations and to conference with peer reviewers. In addition, the teacher provided each student with written feedback. The last component of the baseline data was the ensemble performance video. For this, students performed several warm-ups and two songs which they helped choose. Students were asked to view the video once without sound and respond to what they saw and then they were asked to watch the video again with sound and respond to what they heard. A follow-up discussion also took place.

A larger set of portfolio artifacts was created and collected following the baseline set. This data consisted of more individual and ensemble performance videos along with student evaluations and a collection of reflections. Students prepared two excerpts for their individual performances which were recorded in December. After viewing these performances, students were to fill out individual performance rubrics. Following this, they were to trade tapes and complete peer assessments. The teacher collected and returned copies of these rubrics to students and also provided her own feedback. All document copies stayed in students portfolios. Class time was provided for peer and teacher conferencing. A third round of individual videotaping occurred following the same process in February. The ensemble performance was that of the school's winter concert. Students viewed this video during the class period immediately following the concert and were asked to critique their own performance and the overall ensemble performance. Additional portfolio documents included a student response to a district-

commissioned composer's letter, a reflection following a class unit on vocal pedagogy and adolescent voices, and a reflection on the chorus portfolio process itself.

Data analysis. At the end of the study, portfolio data was sorted and coded according to the following questions:

1. What do students value about chorus?
2. What can students tell us about their own and their peers singing?
3. How has portfolio assessment transformed students' singing from an abstraction to a personal reality over time?
4. How have portfolios helped uncover and foreground student values?
5. How do portfolios reveal student learning as a multi-directional, social, and collaborative process?

Results. It was determined by the researcher that changes in student learning occurred as they "purposefully viewed" their own and their peer's performances, as they "purposefully listened" to their own and their peer's performances, and as they had an opportunity to repeat these tasks over time. Students' visual comments were primarily in regard to the physicality of singing. Many commented on such things as posture, head, chin and mouth positions, facial Expressions, and eye contact.

Regarding listening, students commented on projection, phrasing, and breath support. Many noted that rhythm patterns were easier to perform correctly than tonal ones. They inferred that most rhythm mistakes were caused by incorrect words or missed entrances. Students were not specific when identifying tonal errors although they were more likely to point out an error when an accompanist played the vocal line to assist a student. The terms "sharp, "flat," and "off pitch" were often used incorrectly. A common

concern among students seemed to be with range. Boys showed concern with their changing voices, and girls were concerned with their ability to sing high notes.

The researcher noted that over time, students continued to comment on many of the same aspects as they had initially, but their comments became more detailed and included more pedagogical vocabulary. For example, many students commented not only on their improvement with singing long phrases, but they identified reasons for their improvement such as understanding where appropriate breathing places were located. Also, several students made note of diction clarity, a topic they had not been prompted to include. In addition, boys made mention of being able to sing low notes with more ease and girls focused more on their voice quality than on their upper range, a change from earlier in the study. Two other changes that took place during long-term portfolio use was an increase in student confidence and student acceptance and enjoyment of the portfolio process itself. It was evident at the start of the study that students did not particularly care for portfolio use because of extra work and self-consciousness. One aspect that did not change much during the course of the study was the vagueness and confusion among students in their assessment of tonal accuracy.

Conclusions. McCall (2007) drew several conclusions from this study. First, when chorus students are given multiple opportunities to see and hear themselves perform, they are willing and able to reflect on their solo and ensemble vocal abilities. Also, long-term use of the portfolio process allows students to improve their reflection abilities and to incorporate more music terminology. Finally, since students tend to find it

more difficult to identify aural evidence than visual evidence, more varied assessments are necessary when relying solely on aural evidence.

Misconceptions. The researcher noted several student misconceptions that arose during the course of the study. First, many students used the terms "sharp," "flat," or "off-pitch" to refer to issues with vocal quality rather than to refer to actual pitch problems. Little time was spent on these terms in class and students were not prompted to use them during the course of the study. Also, some of the boys noted improvement in their singing due to an increased volume level and girls explained that taking a bigger breath led to improved vocal quality. Quite a few students seemed to think that rhythmic errors were inevitably caused by incorrect or forgotten words.

Inform teacher learning. As mentioned previously, the researcher was not only interested in how portfolios could inform student learning, but also how they could inform teacher learning. Through the portfolio assessment process, the researcher reported learning about three primary adolescent values (belonging, emotional Expression, and continuity) that entice students to be in music ensembles. Regarding belonging, students listed such reasons for chorus membership as the positive class environment, sharing a common interest in music with fellow students, being with the same students year after year, the allowance to be among friends, and being able to perform as a group as opposed to performing alone. Emotional Expression seems to play a major role in students' interest in chorus. Students wrote of an emotional intensity unique from other classes and the happiness, energy, calmness, and success they felt from

the physical act of singing. Students also wrote of the value of the continuity of their relationship with their teacher and the trust and respect they gained for her.

Comparison to present study. The above study is the only one that has been found involving student self-assessment in a choral setting. It is, however, quite different from the present study on the effects of self-assessment and model-listening on choral performance achievement. First, McCall's (2007) study was strictly qualitative; all data consisted of student writing pieces. Results for the present study consist of final adjudication scores for each of four study groups making it more of a quantitative nature. Also, in McCall's study, choral performance achievement is not a variable like it is in the present study. The purpose of the portfolio study was to see how portfolio use affects student and teacher learning, in an 'academic' sense, whereas the present study was conducted to see how student self-assessment and model-listening affected overall choral achievement. McCall did assess student vocal achievement in her portfolio study for the purposes of providing students with feedback and building a collection of documents for their portfolios. It would have been interesting for McCall to share her findings on student vocal performance change over time with the incorporation of portfolio use. It is certainly important for academic learning to take place in a choral ensemble. It seems, though, that this knowledge is limited in its value if it does not lead to improved musicianship.

Instrumental Self-assessment

Hewitt (2001, 2002) conducted two studies, the first of which is particularly similar to the present study even though its focus is on instrumental music. Hewitt (2001)

conducted a study on the effects of modeling, self-assessment, and self-listening on junior high instrumentalists' music performances and practice attitude. Participants in this study were 82 students from a southwestern state junior high school. These woodwind, brass, and percussion players consisted of 36 seventh graders, 31 eighth graders, and 15 ninth graders. Students were randomly assigned to one of eight treatment groups, each containing either ten or eleven students. These groups contained interactions of two modeling conditions, two self-assessment conditions, and two self-listening conditions.

Materials. Materials needed for this study were a piece of music for students to perform and audiotapes of model performances. The music selected for the study, the Performance Etude, was selected based on specific criteria involving a variety of technical components suitable for junior high musicians and a difficulty level that is appropriate and similar among instrument parts. Three junior high music teachers confirmed the music met the criteria. The model audio-tapes were created by university music majors at a large southwestern university. Professional musicians then selected a recording of each instrumental part to be used as the model.

Method of evaluation. Next, the method of evaluation of student performances and practice attitude was established. Students' performances would be evaluated using the Woodwind Brass Solo Evaluation Form (WBSEF), a five-point rating scale which measures tone, Intonation, technique/articulation, melodic accuracy, Rhythmic Accuracy, tempo, and interpretation. Student practice attitude would be measured using a Practice Attitude Questionnaire (PAQ) created by the researcher. This Likert-type measurement tool would be completed by each student following all treatment sessions.

Procedure. During the first week of this nine-week study, students were trained in the use of the WBSEF. The WBSEF was also piloted and adapted for student self-assessment. The Performance Etude was introduced, with limited instruction, to students this week as well. A performance pre-test was given during the second week in which each student was recorded performing the Practice Etude and asked to practice at home. During the next three weeks, model groups were given model tapes, and self listening groups were given tapes of their own performances. Instructions were provided for each of the eight groups for in-class treatment sessions. The model group, for example, students in Group A (Model X Self-listening X Self-assessment) were instructed to listening to the model recording, perform and record the Performance Etude similarly to the model, listen to their own performance, complete the WBSEF, practice the piece, complete the PAQ, and leave with both the self and model recordings. Students' music performances were assessed for a second time during week eight using the same format used for the pre-test. The last week, 24 students were chosen randomly to be interviewed and videotaped to determine the validity of the PAQ.

Results. It was determined that the 4-way interaction of model, self-listening, self-assessment and test score was not statistically significant. There were, however, two 3-way interactions and one 2-way interaction found to be statistically significant. One was the interaction of model, self-listening, and self-assessment. The other significant 3-way interaction was of the model, self-assessment and test. Follow-up analyses showed statistically significant results for tone, melodic accuracy, Rhythmic Accuracy, interpretation, and Overall Performance scores. It was also shown that the Model X Self-

assessment groups improved more than the No Model X Self-assessment groups for tone, melodic accuracy, Rhythmic Accuracy, interpretation, and Overall Performance. On the other hand, scores did not differ between Model X No Self-assessment groups and No Model X No Self-assessment groups. The statistically significant 2-way interaction was that of model and test. Follow-up analysis showed that all performances areas except Intonation and melodic accuracy had a statistically significant interaction. Students in model groups showed a greater increase sub-areas and overall test scores than students in No-model groups. Student practice attitude scores, which reflected positive attitudes, remained constant throughout the study for all groups.

Model-listening and self-assessment, when done independently, were not shown to have a significant effect on performance achievement. When combined, however, model-listening and self-assessment had a significant effect on performance in the areas of tone, melodic accuracy, Rhythmic Accuracy, interpretation, and Overall Performance. Self-listening did not have a significant effect on performance achievement.

Comparison to present study. As in the present study, Hewitt was interested in the effects of model-listening and self-assessment on middle school ensemble performance achievement. Hewitt's study was a little more complex, however, in that it tied in a third independent variable: self-listening. Considering self-listening did not warrant any significant results in Hewitt's study, and in an effort to simplify, self-listening will be considered a necessary component of self-assessment in the present study. Another aspect of Hewitt's study which was not incorporated into the present study was the analyses of student practice attitude. This was important in Hewitt's study

since band students, unlike choral students, are typically required to practice on a regular basis outside of school. A practice method, as effective as it may be in the short term, holds little value if students are disinterested and unwilling to continue its use. It was determined that the aspect of practice attitude, in a choral study, was less necessary and therefore would not be included.

Hewitt (2002) conducted a similar study on self-assessment tendencies of junior high instrumentalists. Hewitt was interested to see how student self-assessments evolved over time, whether the process of self-assessment with or without a model has an effect on self-assessment accuracy, and to see if there is a relationship between self-assessment accuracy and music performance. Participants in this pretest/posttest study were 41 junior high woodwind, brass, and percussion students in grades seven and eight. Students were auditioned and placed into one of three groups: low ability, middle ability, and high ability. Students were randomly assigned to a group with the presence or absence of an aural model.

Materials. Materials for this study were a performance assessment (the Woodwind Brass Solo Evaluation Form), an adapted version of the WBSEF (the result of a pilot session) for student self-assessments, performance music, and model recordings. The music selected for the study, three short etudes, was selected based on specific criteria involving a variety of technical components suitable for junior high musicians, and a difficulty level that is appropriate and similar among instrument parts. Three junior high music teachers confirmed that the music did meet the criteria. The model audio-

tapes were created by university music major. Professional musicians then selected a recording of each instrumental part to be used as the model.

Procedure. During week one of this six-week study, students were trained to use the adapted version of the WBSEF. Also, students had an opportunity to play through the etudes as a group. Students recorded and evaluated their individual performances during week two. They were also asked to practice the music outside of school. The next three weeks were designated for weekly performances and self-assessments. Those in the model treatment groups were also instructed to listen to the model recording prior to each performance and also whenever they practiced at home. Students recorded their performances the last week and self-evaluated.

Results. Hewitt found that while student self-assessment scores did increase over time, they did not increase in accuracy. The presence or absence of an aural model had no bearing on this result. Self-assessment accuracy was actually found to decrease over time in the area of Intonation. Furthermore, correlations were low for music performance and self-assessment accuracy.

Comparison to present study. It is curious that Hewitt's study, on the evolvement of student self-assessment, was conducted during such a short period of time: only six weeks. Perhaps Hewitt's lack of positive findings was due in part to the fact that students were not given a reasonable amount of time to show signs of improvement. It also seems that more focus should have been given to the process of training students to self-assess and that more independent variables (not just an aural model) should have been incorporated. The duration of the present study was only three weeks longer than

Hewitt's and the self-assessment training process involved a comparable amount of rigor. However, the present study's focus was not on the *evolvement* of student self-assessment, but on the effect of self-assessment on performance achievement.

Model-listening

A study was done by Morrison and Montemayor (2004) on the effect of a recorded model on band students' performance self-assessments, achievement, and attitude. Participants for this five-week study were band students from three middle/junior high schools and two high schools. The schools spanned urban, suburban, and rural areas in the Pacific Northwest and had well-established instrumental programs.

Materials. Materials needed for this five-week study were performances pieces, model recordings of the performances pieces, student progress reports, and a method of final performance evaluation. The performance pieces consisted of two single-movement concert works per school which were chosen by the five participating directors based on difficulty level. Professional or collegiate compact disc recordings of one of the two performance pieces were distributed by the researchers to each school. The student progress report contained five questions, four of which were free response and a fifth which contained eight parts with Likert-scale responses. On this progress report, students were asked such things as what they learned in a given day, or what aspects of their performance were strong. The fifth question asked students how much they liked the piece so far which provided evidence of the effect of model-use on student attitude. For final performance evaluation, five experienced instrumental music teachers evaluated

each excerpt in the areas of notes/rhythms, articulation/dynamics, Intonation, and balance using a 5-point Likert-type scale.

Procedure. During the first week, pre-treatment ensemble recordings were made of both pieces in each school. Following this, the directors spent an equal amount of rehearsal time on both pieces. The directors also played the model recording in its entirety once a week while band members followed along in their music. On a different day each week, directors play a small portion of the model recording aligning with what was rehearsed that day. Students completed the progress reports each week on a day that they rehearsed the pieces. This would be a separate progress report for each piece, each completed on a separate day of the week. Post-treatment ensemble recordings were made the final week of the study.

Data analysis. ANOVA was used to compare the mean pre-treatment and post-treatment evaluations on each of the four criteria. The result was that the post-treatment scores were significantly higher than pre-treatment scores for all four criteria regardless of level (middle/junior or high school) or presence of an aural model. Results using repeated measure multivariate ANOVA showed that high school students' evaluations were significantly greater than middle/junior school students between week one and week five. In a comparison of student self-assessments of individual performance with self-assessments of ensemble performance according to condition (model or no model) using multivariate ANOVA, it was found that high school students consistently rated no-model performances higher than model performances. Also, high school students were consistently more positive in rating their own performance versus that of the ensemble.

Repeated-measures ANOVA was used to compare students' responses to the fifth question on the progress report concerning attitude according to model and no-model situations. Middle/junior high school students showed a significantly stronger response to pieces that were learned with help of an aural model whereas high school students did not seem to have a preference.

Comparison. Morrison and Montemayor (2004) examined the effects of an aural model on middle school ensemble students' self-assessment accuracy and on their performance achievement. This is similar to the present study except that here, self-assessment was considered a dependent variable along with performance achievement. Also, Morrison and Montemayor (2004) factor in the aspect of practice attitude. As in Hewitt's (2001) study, this is well-fitting since the study is concerned with instrumentalists, not choral singers.

Summary

The only similarities between all of the above studies and the present one were the participant age group (middle school) and the inclusion of self-assessment. Even though self-assessment was a factor in all four studies, it was only used as an independent variable, as it was in the present study, in the two studies done by Hewitt. These two studies are also the only ones which examined the effects of self-assessment on performance achievement. The three instrumental studies are similar to the study at hand in that they all examined the effect of an aural model on performance achievement.

One important difference between the reviewed studies and the present one is that all reviewed studies involved assessment (either by students or the researcher) of

individual musicians. The present study investigated assessment only of music ensembles.

CHAPTER 3

METHODOLOGY

Introduction

Research indicates that self-assessment in a choral ensemble leads to: (a) increased motivation, (b) better understanding, (c) improved student-to-teacher communication, (d) time utilization, and (e) the promotion of independent musicianship. However, research has not indicated a correlation between self-assessment and choral performance achievement. Therefore, the present study was concerned with the effect of the following three conditions: self-assessment, model-listening, and self-assessment combined with model-listening on choral performance achievement. The purpose of this study was to investigate the following questions concerning a middle school choral ensemble:

1. Does routine student self-assessment improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

2. Does routine model-listening improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

3. Does the combination of routine model-listening and self-assessment improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

Participants

This study was implemented over a 10-week period with members of a suburban charter school choral ensemble. This ensemble, the Concert Choir, is a 6-8th grade non-audition 3-part mixed chorus of 64 students with varying levels of experience. Students in grades K-8 attend this suburban charter school and the middle school (grades 5-8) has an enrollment of approximately 700 students. The male to female ratio is 1:1 and the socio-economic status among student families is quite varied. The school draws its population of students from an area within a five-mile radius of the school: an area that encompasses low, middle, and high-income families in both rural and urban settings. The Concert Choir rehearses one day per week after school for an hour and fifteen minutes. Students also attend rotating sectional (soprano, alto, or tenor) rehearsals on Wednesdays and Thursdays during recess for about 25 minutes.

An informational letter was sent home to parents of study participants prior to the start of the study (See Appendix C) and the Protection of Human Subjects course has been completed (See Appendix D). Also, permission has been granted to conduct research involving human subjects by the Office of the Vice Provost for Research (See Appendix E).

Procedure

To investigate the above research questions, students were randomly assigned to one of four groups (three treatment groups and one no-treatment group) with similar balance of voices. Similar to Hewitt's (2001) study on the effects of modeling, self-assessment, and self-listening on junior high instrumentalists' music performance and

practice attitude, conditions for this study were as follows: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-Assessment.

Throughout the course of the study, the Concert Choir had their regularly-scheduled whole-group rehearsals for an hour and fifteen minutes once per week after school. During this time, the study repertoire was rehearsed in exactly the same manner as the other performance repertoire. The other music had no bearing on the study. Each group met once per week during recess to self-assess and/or listen to the model CD. Students in the no model/no self-assessment group were asked to complete choral worksheets from the book "A Workbook in the Fundamentals of Singing" (Swift, 1958).

Week 1. According to Robinson (1990) student self-assessment accuracy is likely to improve with training. Therefore, during week one, to increase the likelihood of accurate student self-assessment, the two self-assessment groups received training for 45 minutes. For the training process, choral rating scales terms and completion instructions were briefly discussed. Next, students listened to five choral recordings and completed the Student Choral Rating Scale (See Appendix B) for each. After each listening/assessment, student responses were discussed and students were encouraged to verbally share their responses.

Weeks two through nine. During the next eight weeks, self-assessment groups made recordings and completed self-assessments on a weekly basis. For each of these sessions, a self-assessment group sang a designated portion of the choral piece chosen for the study. The portion of music was determined each week based on what was covered

during the "normal ensemble rehearsal." Performances were recorded on a hand-held digital tape recorder. The group listened to the entire recording and then each student in the group was asked to complete a Student Choral Rating Scale on which they rated their ensemble's performance achievement. Model-listening groups attended weekly sessions that were scheduled during recess to listen to the model CD (Gray, 2009). The recording was of a choir performing the piece selected for the study that excelled in the performance sub-areas indicated on the Student Choral Rating Scale. The weekly sessions lasted about 25 minutes.

Week ten. During week ten, each of the four groups made a final recording. These recorded final performances were evaluated using the Teacher Choral Rating Scale (See Appendix A) by two independent judges.

Preparation of Materials

Materials necessary for this study were: (a) a piece of choral music and the criteria for its selection, (b) a recording of a model choir performing the selected piece of music for model-listening purposes, (c) a recording of five middle school choral performances of varying levels for pilot-testing of the choral rating scale and for training participants to self-assess ensemble achievement, and (d) a means for recording students' performances for weekly self-assessment and final adjudication purposes.

The musical selection to be used for this study was Cynthia Gray's "Song of the Wind" (2009) published by Lorenz Publishing Company. The factors considered in choosing this piece were its 3-part mixed voicing, easy-medium difficulty level, the need for clear diction and vowel purity, and its variation in style, articulation and dynamic

level. An easy-medium piece of repertoire was chosen to enable a measurable amount of achievement during this short, 10-week study. The piece's demand for clear diction and vowel purity did, in part, allow for assessment of Tone Quality, Intonation, and Balance/Blend; its variation in style helped accommodate assessment of Expression. The recording of a model choir was a performance CD of "Song of the Wind," also a product of the Lorenz Publishing Company (2009).

The following five choral recordings used for the self-assessment pilot test and training were chosen for their variety in choral performance quality:

1. Patterson (2009) This Shall Be for Music, performed by the Arthurs Middle School 7th Grade Chorus from Ypsilanti, Michigan

2. Copland (2009) Ching a Ring Chaw, performed by the Arthurs Middle School 7th Grade Chorus from Ypsilanti, Michigan

3. Shaw (2009) Rock a My Soul, arrange, performed by the Port Jervis Middle School 7th Grade Chorus from Port Jervis, New York

4. Thomas (1999) Keep Your Lamps, performed by the 1999 Delaware Junior All-State Chorus

5. Emerson (1999) Shoshone Love Song, performed by the 1999 Delaware Junior All-State Chorus

All choral recordings were made using the Zoom H2 Portable 2-Track SD Recorder made by Samson.

Data Collection

Methods for collecting data were: (a) Teacher Choral Rating Scale (See Appendix A) to measure the choral achievement for each of the four groups at the end of the study and (b) Student Choral Rating Scale (See Appendix B) used by the two self-assessment groups once per week.

The Teacher Choral Rating Scale used in this study was "The Five Dimensions of Achievement in Choral Music Performances" (Larkin, 1985) (See Appendix A). This assessment tool was used at the end of the study by two independent judges to adjudicate a final performance of each of the four choral groups. To assure consistency in final scores, the independent judges completed sample Teacher Choral Rating Scales in the same manner as the students, with the same materials. The judges went over their scores and discussed discrepancies. This piloting session resulted in two changes to the Teacher Choral Rating Scale: the incorporation of a numeric point system and the removal of unclear or redundant criteria. The adapted Teacher Choral Rating Scale was used by two independent judges to score each of the four study groups in each of the five performance sub-areas (Tone Quality, etc).

The Teacher Choral Rating Scale was adapted for student use and piloted on a small group of choral students. For this process, students used the Teacher Choral Rating while being asked questions about terms, instructions and other points of clarity regarding the rating scale. The pilot study participants listened to the same five recordings used to train study participants to self-assess. They were asked to listen to each recording and to complete the Teacher Choral Rating Scale. After each listening/assessment, student responses were discussed. Findings from the pilot test determined necessary adaptations to increase reliability and validity of the rating scale. The resulting Student Choral Rating Scale (See Appendix B) was used weekly by self-assessment groups. These were collected and compared with teacher-completed choral rating scales to test for accuracy in student self-assessment.

Data Analysis

Data consisted of final choral ratings for each of the four conditions: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment. Data were analyzed using descriptive statistics to determine differences among them for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement.

Interjudge Reliability

Final choral performance scores were rated by two independent judges. Interjudge reliabilities were determined for each performance sub-area and for Overall Performance using a Pearson-product moment correlation.

Validity

The validity of this study was ensured in several ways: (a) students were randomly assigned to treatment groups, (b) a rationale was provided for the musical selection, (c) the Student Choral Rating Scale was piloted on choral students, (d) self-assessment groups received training in the use of the Student Choral Rating Scale, (e) the Teacher Choral Rating Scale was piloted on judges, and (f) two independent judges assessed final recordings.

Summary

For this study, 64 middle school chorus students were assigned to one of the following four groups: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment. Once per

week, self-assessment groups completed self-assessments and model-listening groups listened to a model recording. During week ten, each group made final recordings which were evaluated by two judges. Data were analyzed using descriptive statistics to determine differences among the groups for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement. Data analysis led to answers for each of the research questions.

CHAPTER 4

DATA ANALYSIS AND RESULTS

The purpose of this study was to determine the effects of routine self-assessment and model-listening on choral achievement. The research questions concerning assessment in a middle school choral ensemble are as follows:

1. Does routine student self-assessment improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

2. Does routine model-listening improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

3. Does the combination of routine model-listening and self-assessment improve performance achievement in the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

Sixty-four middle school chorus students were assigned to one of the following four groups: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment. Once per week, self-assessment groups completed self-assessments and model-listening groups listened to a model recording. During week ten, each group made final recordings which were evaluated by two judges.

To examine these questions, data were collected and evaluated using the following statistics: (a) interjudge reliabilities for the Teacher Choral Rating Scales, and

(b) the means and standard deviations of the choral scores. Two independent judges rated each group's final performance using the Teacher Choral Rating Scale.

Interjudge Reliability

Final choral performance scores were rated by two independent judges using the Teacher Choral Rating Scale (see Appendix A). Interjudge reliabilities were determined for each performance sub-area and for Overall Performance using Pearson-product moment correlations.

As shown in Table 1, interjudge reliabilities ranged from 0.246-0.926. Reliabilities were moderately weak for Rhythmic Accuracy, moderately strong for Expression, and quite strong for Tone Quality, Intonation, Balance/Blend, and Overall Performance.

Table 1

Interjudge Reliabilities

Performance Sub-area	*r*
Tone Quality	0.873
Intonation	0.926
Rhythmic Accuracy	0.246
Expression	0.662
Balance/Blend	0.923
Overall Performance	0.794

Data Analysis

Data consisting of mean choral ratings for each of the four conditions: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-Assessment were graphed as bar graphs (see Figures 1-2). Data were computed for the mean and standard deviation of the judges' scores for each sub-area (see Tables 2-3). The mean was used as a baseline to compare differences among the groups for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement.

Results

Research question one. To address research question one, concerning the effect of student self-assessment on choral performance achievement, mean self-assessment group scores were compared with mean no-treatment group scores for each performance sub-area. Table 2 shows means and standard deviations of judges' scores for the self-assessment group and the no-treatment group in each performance sub-area.

Table 2

Means and standard deviations of self-assessment and no-treatment group scores

	No Model X Self-assessment		**No Model X No Self-assessment**	
	M	SD	M	SD
Tone Quality	8.500	2.121	10.000	1.414
Intonation	8.500	2.121	10.500	2.121
Rhythmic Accuracy	8.500	2.121	10.500	2.121
Expression	10.000	1.414	11.000	2.828
Balance/ Blend	9.000	2.828	6.000	2.828

As shown in Figure 1, self-assessment alone was more effective than no-treatment in the performance sub-area of Balance/Blend. The no-treatment group received higher mean scores. No-treatment was the most effective for Tone Quality, Intonation, and Expression. No treatment was also the most effective for Rhythmic Accuracy, for which it should be noted that interjudge reliability was low making this result questionable.

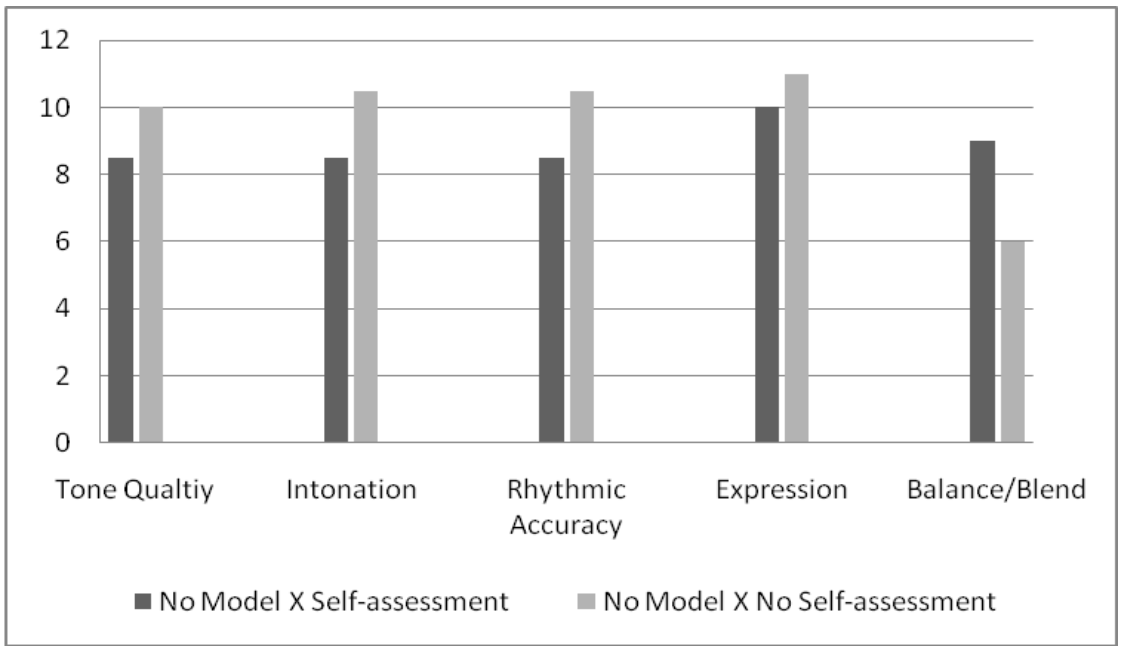

Figure 1: Comparison of self-assessment and no-treatment mean group scores

Research question two. To address research question two, which concerns the effect of model-listening on choral performance achievement, mean model group scores were compared with mean no-treatment group scores for each performance sub-area. Means and standard deviations of judges' scores are shown in Table 3 for the model-listening group and the no-treatment group in each performance sub-area.

Table 3

Means and standard deviations of model and no-treatment group scores

	Model X No Self-assessment		**No Model X No Self-assessment**	
	M	SD	M	SD
Tone Quality	6.000	0	10.000	1.414
Intonation	5.000	1.414	10.500	2.121
Rhythmic Accuracy	8.500	0.707	10.500	2.121
Expression	9.000	1.414	11.000	2.828
Balance/ Blend	7.500	2.121	6.000	2.828

As displayed in Figure 2, model-listening alone led to slightly higher means than no treatment in the performance sub-area of Balance/Blend. The no-treatment group had the highest means for Tone Quality, Intonation, and Expression. No-treatment also led to the highest mean for Rhythmic Accuracy although low interjudge reliability must be considered here.

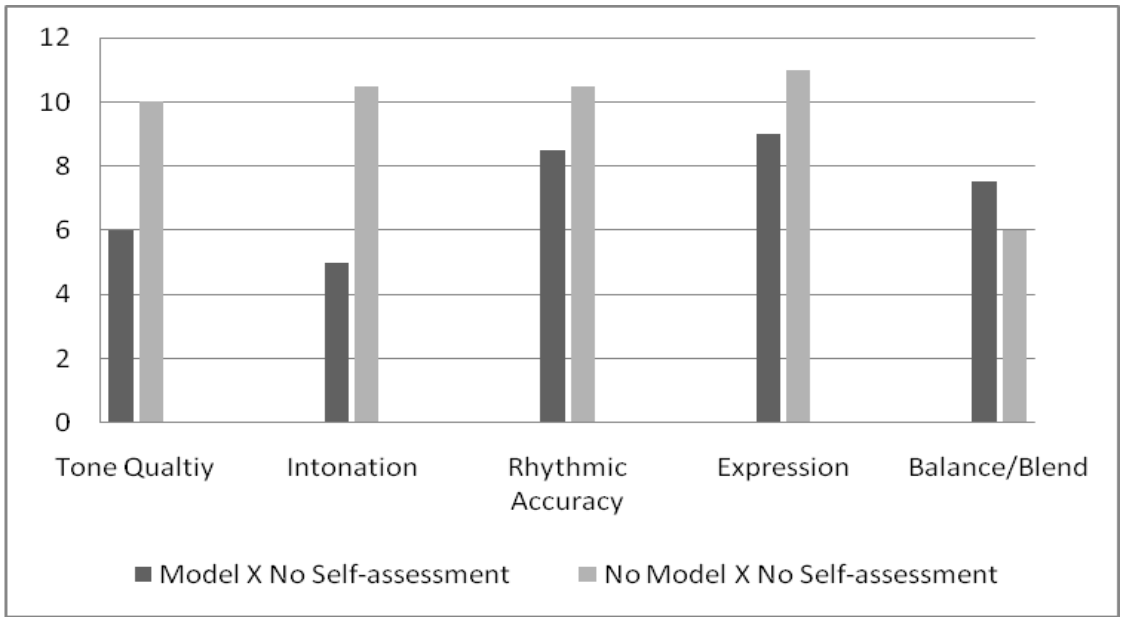

Figure 2: Comparison of model and no-treatment mean group scores

Research question three. To address research question three, concerning the effect of student self-assessment combined with model-listening on total choral performance achievement, mean scores of self-assessment only, model-listening only, self-assessment combined with model-listening and no-treatment groups were compared for each performance sub-area. Table 4 shows the Overall Performance means and standard deviations for each group and performance sub-area.

Table 4

Overall Performance Means and Standard Deviations

	Model X Self-assessment		**Model X No Self-assessment**		**No Model X Self-assessment**		**No Model X No Self-assessment**	
	M	SD	M	SD	M	SD	M	SD
Tone Quality	9.000	1.414	6.000	0	8.500	2.121	**10.000**	1.414
Intonation	8.500	0.707	5.000	1.414	8.500	2.121	**10.500**	2.121
Rhythmic Accuracy	9.000	0	8.500	0.707	8.500	2.121	**10.500**	2.121
Expression	10.000	1.414	9.000	1.414	10.000	1.414	**11.000**	2.828
Balance/ Blend	8.000	2.828	7.500	2.121	**9.000**	2.828	6.000	2.828
Overall Performance	45.000	5.657	36.000	5.657	44.500	10.607	**48.000**	11.314

Figure 3 illustrates that no-treatment was most effective in developing the sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Overall Performance. Self-assessment alone was most effective for the development of Balance and Blend.

The no-treatment group had the highest mean for tone quality. However, self-assessment combined with model listening had the lowest mean, suggesting that self-assessment and model-listening may hinder tone quality development.

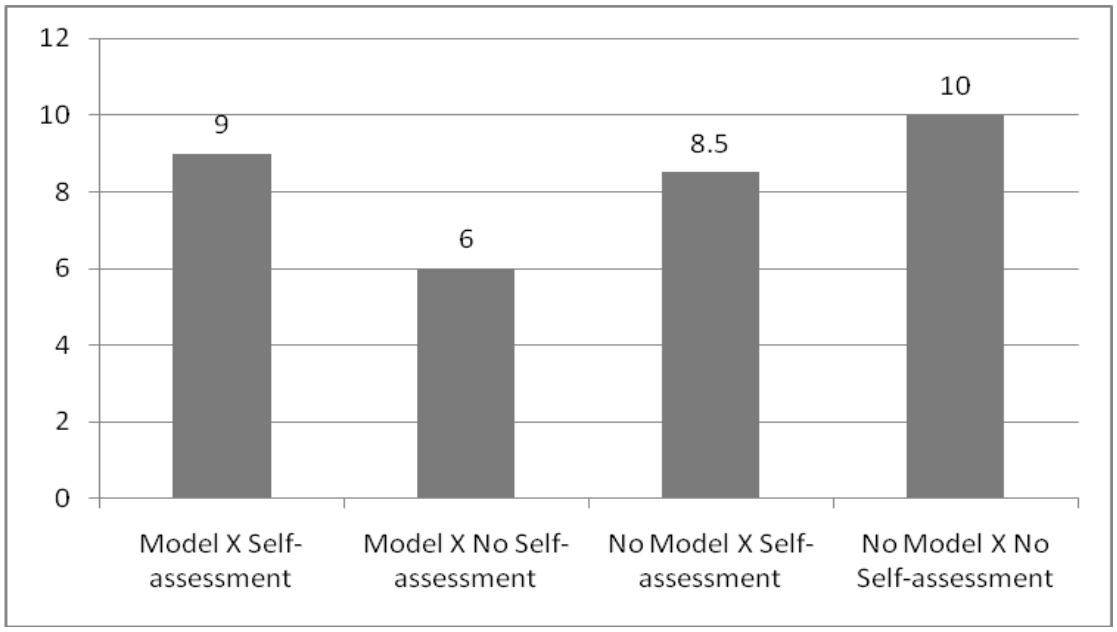

Figure 3: Comparison of mean group scores for Tone Quality

The no-treatment group had the highest mean for Intonation. Self-assessment combined with model-listening was ineffective in improving choral performance achievement in the performance sub-area of Intonation as shown in Figure 4.

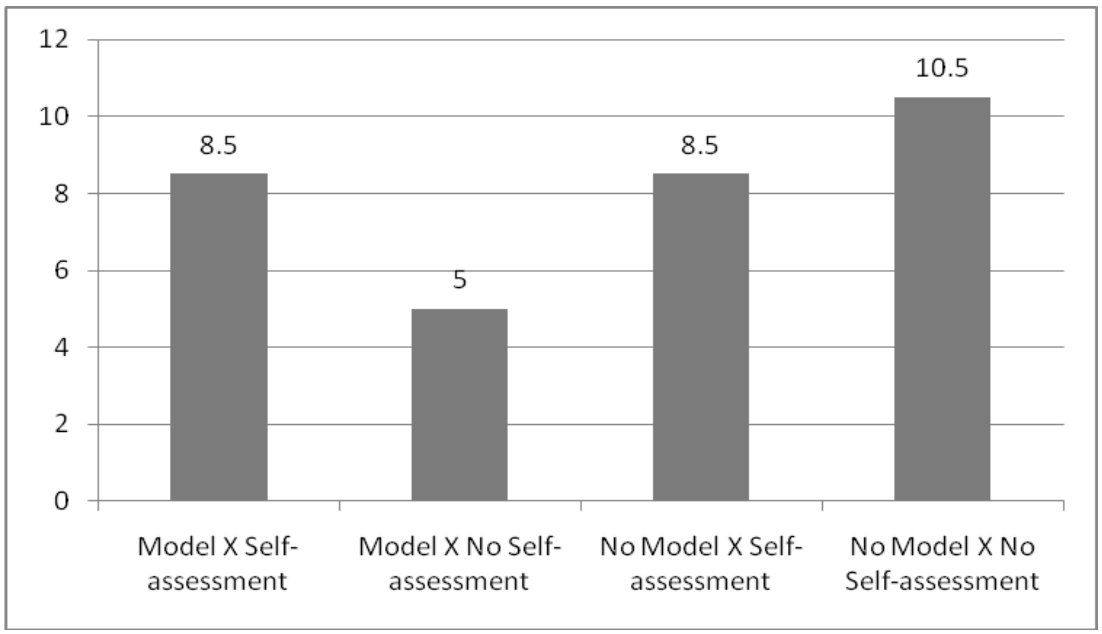

Figure 4: Comparison of mean group scores for Intonation

Figure 5 shows the no-treatment group had the highest mean for the sub-area of Rhythmic Accuracy. It should be noted that interjudge reliability was low for Rhythmic Accuracy making these results questionable.

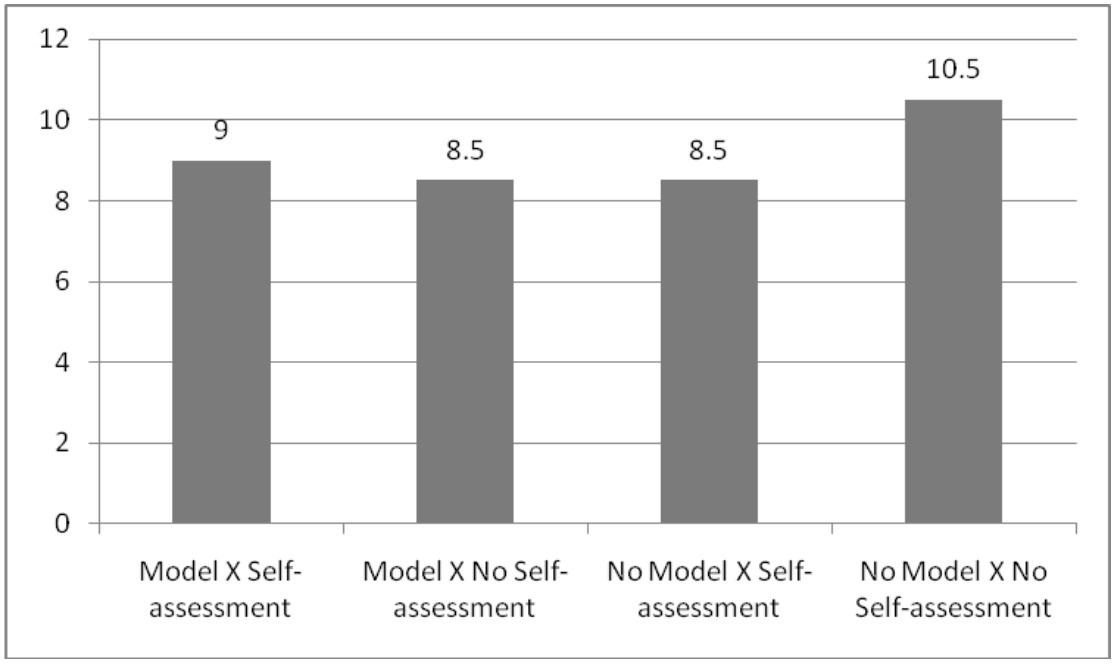

Figure 5: Comparison of mean group scores for Rhythmic Accuracy

As shown in Figure 6, the no-treatment group had the highest mean for the performance sub-area of Expression. Model-listening alone led to the lowest Expression score.

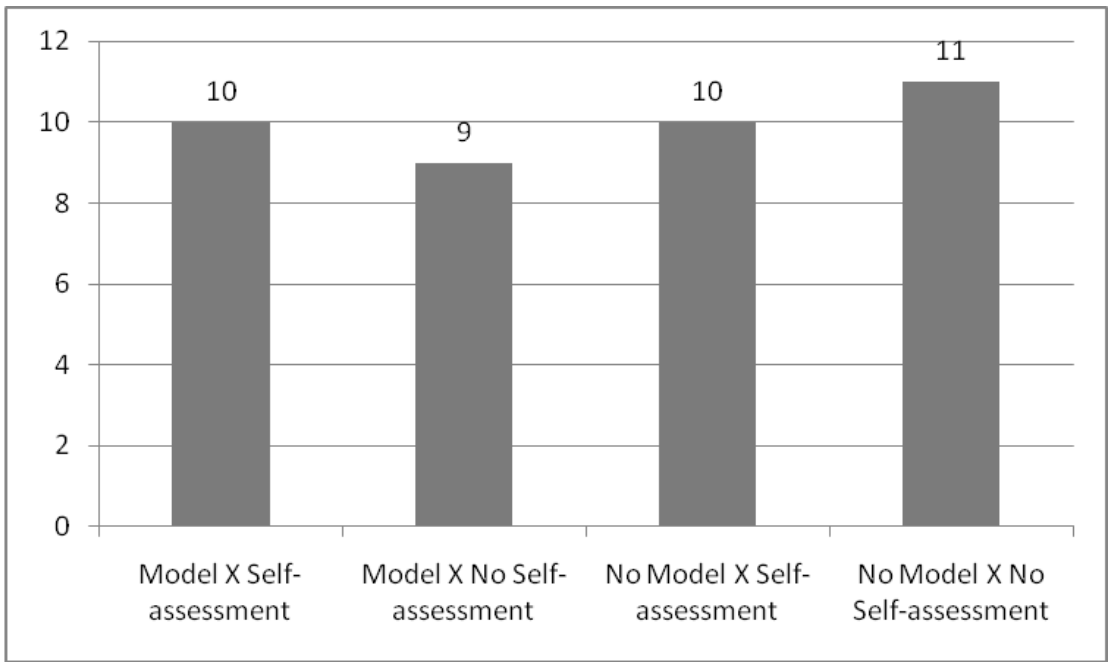

Figure 6: Comparison of mean group scores for Expression

The self-assessment group had the highest mean for Balance/Blend, indicating that self-assessment was the most effective in improving Balance/Blend (see Figure 7).

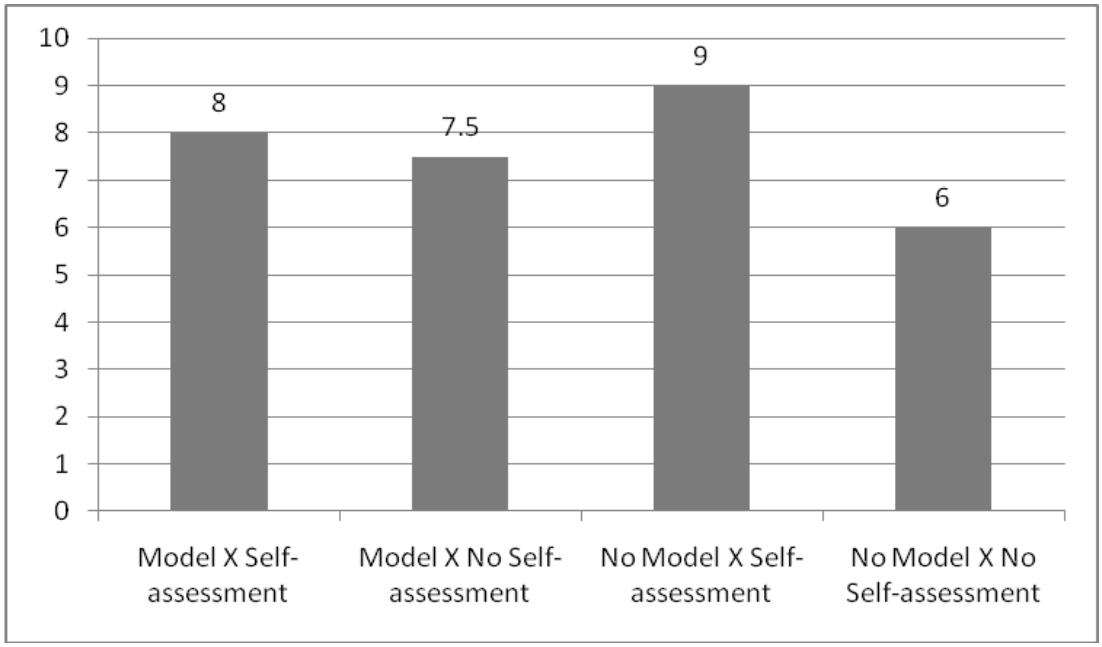

Figure 7: Comparison of mean group scores for Balance/Blend

As can be seen in Figure 8, the no-treatment group had the highest Overall Performance score. Self-assessment with and without model-listening produced comparable results to each other although less were effective than no-treatment. Model-listening led to a particularly low Overall Performance score.

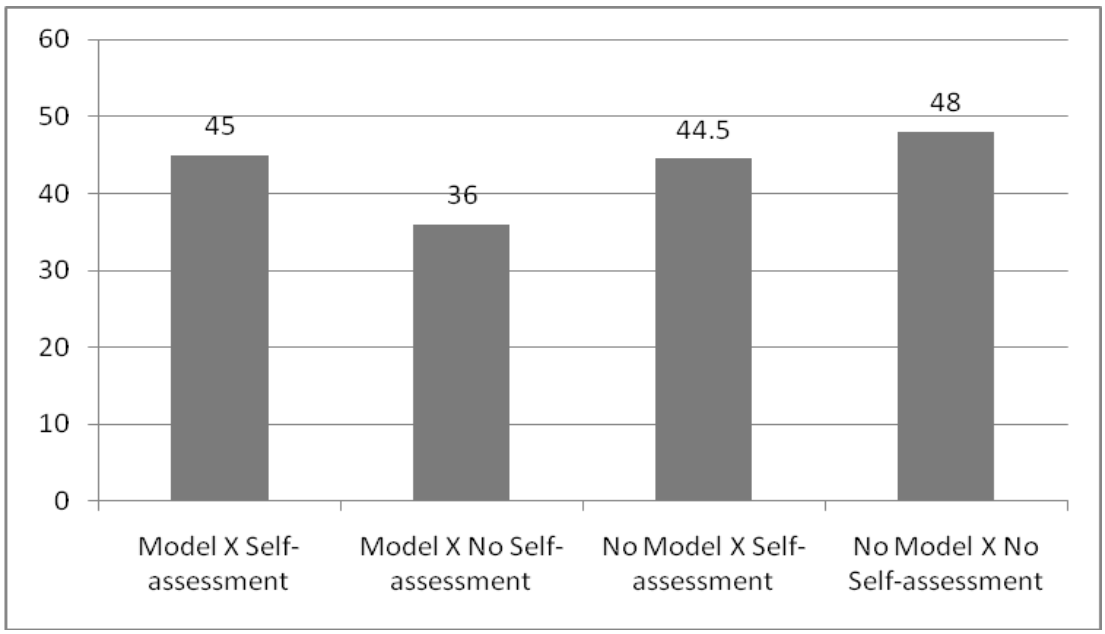

Figure 8: Comparison of mean group scores for Overall Performance

Summary

At the end of the study, criterion performances were evaluated by two judges. Data were analyzed using descriptive statistics to determine differences among the groups for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement. The mean was used as a baseline to compare differences among the groups for the five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement.

According to the results of the present study, the no-treatment group scored the highest in the performance sub-areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and in Overall Performance. Self-assessment led to improved performance

achievement in the sub-area of Balance/Blend. Model-listening led to the lowest score for Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Overall Performance. Tone Quality and Intonation means were particularly low for the model-listening group.

There was a notable trend for each performance sub-area and for Overall Performance among the four study groups. Typically, the no-treatment group scored the highest, the model/no self-assessment and no model/self-assessment groups had similar middle-range scores, and the model/no self-assessment group scored the lowest on the criterion performance measure.

These results provide insight into the effects of self-assessment and model-listening on middle school choral performance achievement. Also, these results address the following questions concerning assessment in a middle school choral ensemble:

1. Does routine student self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

2. Does routine model-listening improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

3. Does the combination of routine model-listening and self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

In Chapter 5, conclusions, implications for practice in the choral classroom, and suggestions for future research will be given.

CHAPTER 5

CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS FOR FUTURE RESEARCH

Summary

Student self-assessment in the choral music classroom is a topic that has been insufficiently addressed by researchers. The research literature demonstrates the positive effects that student self-assessment can have on instrumental ensembles and on learning in other classroom settings (Matheny, 1994, McCall, 2007). The present study was conducted to provide the music education profession with more information on student self-assessment in the choral spectrum.

The purpose of this study was to determine the effects of routine self-assessment and model-listening on choral achievement.

Specifically, the research questions, concerning a middle school chorus, were:

1. Does routine student self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

2. Does routine model-listening improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

3. Does the combination of routine model-listening and self-assessment improve performance achievement in the areas of Tone Quality, Intonation, Rhythmic Accuracy, Expression, and/or Balance/Blend?

The study was implemented over a 10-week period with members of a suburban charter school choral ensemble. The ensemble, the Concert Choir, was a 6-8th grade non-auditioned, 3-part mixed chorus of 64 students with varying levels of musical experience. Students were randomly assigned to one of four groups (three treatment groups and one control group) with similar balances of voice. Conditions for this study were as follows: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment.

During week one of this ten-week study, the two self-assessment groups received training. During weeks two through nine, the self-assessment groups recorded performances and completed self-assessments on a weekly basis. Also during this time, model-listening groups attended weekly meetings to listen to a model recording. During week ten, each of the four groups made final recordings of "Song of the Wind" (2009) by Cynthia Gray. These recordings were evaluated by two independent judges.

Data consisted of final choral ratings for each of the four conditions: (a) Model X Self-Assessment, (b) Model X No Self-Assessment, (c) No Model X Self-Assessment, and (d) No Model X No Self-assessment. To determine the effect of self-assessment and model-listening on choral performance achievement, group performance scores were compared for each of five performance sub-areas (Tone Quality, Intonation, Rhythmic Accuracy, Expression, and Balance/Blend) and for Overall Performance achievement.

Delimitations of the Study

Some aspects of the methodology in this study may have caused the results to lack accuracy. The aspect of most concern was that only one result was produced for each

treatment. This is because each of the four groups was assessed as a collective ensemble. Even though students were randomly assigned to treatment groups, the likelihood exists that one group might possess a higher performance achievement level regardless of imposed treatment. To remedy this problem in future studies, the number of groups per treatment must be increased. Participants should be randomly selected for group placement as they were in the present study to ensure equal ability levels among groups.

Another aspect which may have impacted the accuracy of the results was that adolescent students may have felt uncomfortable in small self-assessment groups. Each group had only 16 students; it was apparent during recording sessions that a number of students were inhibited because they were embarrassed that they would be heard individually on the recording.

Another delimitation of this study was that the groups met during their recess time. While there were consequences in place for students who missed the research study sessions, many instances existed in which students chose to go to recess instead. Also, many students had excused absences from study sessions due to such obligations as test make-ups or help sessions with other teachers of other curricular areas.

One other possible delimitation was that the study duration was only 10 weeks long with groups meeting once per week for 30 minutes. This meant that students spent a total of five hours in groups during the entire study. The five hours of instruction included one 45-minute self-assessment training session and time for recording final performances. The amount of time devoted to any of the treatment options, or no-

treatment, and training students to self-assess their musical performance was likely insufficient.

Results

According to the results of this study, the no-treatment group scored the highest in the performance sub-areas of Tone Quality, Intonation, Expression, and Rhythmic Accuracy, and in Overall Performance. Self-assessment yielded the highest Balance/Blend score. Model-listening alone appeared to have a negative impact on the performance sub-areas of Tone Quality and Intonation.

Conclusions

From the results of this study, the following conclusions may be made. These conclusions are not generalizable due to the small sample size and short study duration. They may be transferable to other settings however caution is warranted in extrapolating these conclusions beyond the study population.

Conclusion one. Middle school choral students in this study benefited the most from no-treatment in the performance sub-areas of Tone Quality, Intonation, and Expression, and in Overall Performance. These particular performance sub-areas may be difficult for middle school students to comprehend, in an academic sense, particularly with insufficient time to teach students these concepts. Self-assessment and model-listening may only lead to further confusion of how to achieve good Tone Quality, Intonation, and Expression. Perhaps these sub-areas would have been improved by the director teaching the concepts, leading students to listen to models for these musical

nuances, and providing conducting gestures and demonstrations that assist the choir in these sub-areas.

Conclusion two. Model-listening alone may be detrimental to middle school choral performance achievement, particularly in the sub-areas of Tone Quality and Intonation. This conclusion conflicts with Hewitt's (2001) finding that model-listening increases middle school instrumental performance achievement in the sub-area of tone quality. On the other hand, Hewitt (2001) found that model-listening had no effect on intonation. Morrison, Montemayor, & Wiltshire (2004) also determined that model-listening had no effect on secondary instrumental performance achievement in any sub-area. Perhaps, in the present study, participants imitated aural models to the extent that their resulting performance was unauthentic. Another possibility is that students' confidence decreased upon listening to a model which they felt was beyond their performance ability.

Conclusion three. Self-assessment alone may be the most effective treatment for improving middle school choral performance achievement in the performance sub-area of Balance/Blend. Students in the present study seemed to take a particular interest in listening for individuals or voice parts to 'stick' out when completing the rating scales. This makes sense given the self-conscious nature of the adolescent student participants.

Conclusion four. Self-assessment and model-listening may cause adolescent choral students to lose confidence in their singing abilities. Teenagers are at a vulnerable stage of development and need a great deal of encouragement and positive reinforcement. Self-assessment, especially if implemented in a choir's early stages of development,

could cause students to be overly critical and lead to a decrease in confidence. Model-listening has the potential to present students with a performance that is *ideal*, thus leading students to believe that they are incapable of achieving at this level.

Implications for Music Education

Even though the present study suggested limited benefits of self-assessment for choral performance achievement, choral and instrumental educators should consider its use in the classroom. Some benefits of self-assessment in the choral ensemble demonstrated in research literature are that student motivation is increased and student-director communication is improved (Matheny, 1994, McCall, 2007). One primary goal for educators, no matter the subject area, is to empower students to be independent learners. Students must learn for themselves and apply their knowledge in order for it to be meaningful and long-lasting (Wiggins & McTighe, 2006). Therefore, students who participate in music ensembles must be given as much opportunity as possible to be involved in rehearsal decision-making. Student involvement in the assessment process is likely to be a substantial part of achieving this goal.

There are many ways of implementing self-assessment into a choral curriculum. Teachers may decide to include self-assessment by having students routinely complete the Student Choral Rating Scale, or something comparable, as was done in the present study. There are other options as well. Teachers may choose a more formal approach, such as portfolio assessment, as McCall (2007) did in her study with extensive portfolio use. An informal approach such as open-ended class discussion may be just as beneficial

as the above two methods. All choral programs and directors are unique and will likely benefit from different methods of self-and group-assessment.

Suggestions for Future Research

The questions posed in this study are important for several reasons. First, student self-assessment in choral music is a topic that has not been given enough attention in the music research field. Also, the number of documented studies is relatively small with the majority of studies pertaining specifically to instrumental ensembles. Previous studies have shown the positive effects that student self-assessment can have on instrumental ensemble achievement and on achievement in many other curricular areas. The music education profession now needs more information on the effects of this teaching tool in the choral spectrum.

There are several possibilities for future expansion or alteration of the present study. Hewitt (2001) and Morrison, Montemayor and Wiltshire (2004) incorporated measuring students' practice attitude into their instrumental studies. Practice attitude was not considered in the present choral study because choral students do not typically "practice" outside rehearsals to the same extent as instrumental students. However, it may be a worthwhile endeavor to measure student attitude toward self-assessing and listening to a model *during* rehearsals. This could be enacted by having students complete rehearsal attitude surveys as they did in the studies mentioned above. Whether or not self-assessment and model-listening have positive effects on students' musical achievement, their values are limited if they cause students' rehearsal enjoyment to decline

Future studies may narrow the amount of performance sub-areas studied. The present study suggested that Balance and Blend was the only performance sub-area which was positively affected by self-assessment; future researchers may choose to focus on that particular sub-area. Rhythmic Accuracy was a performance sub-area which seemed unaffected by self-assessment and model-listening and might also be worth focusing on in future research.

Another adaption that could be made to the methodology of the present study is for the participants to be elementary or high school students instead of middle school students. The results may be dramatically different with these two age groups since older students would probably be less self-conscious. Other ways to lessen the factor of self-consciousness would be to increase group sizes or to conduct the study over a longer time frame so that students could get used to being recorded.

Model-listening, as done in the present study, was ineffective for some choral performance sub-areas and detrimental for others. Researchers should document the effect model-listening has *over time* on middle school students' choral performance. Additionally, middle school chorus teachers should consider implementing self-assessment into their curriculum. The present study suggested a positive correlation between self-assessment and performance achievement in the sub-area of Balance/Blend. While self-assessment did not seem to benefit performance achievement in any other sub-area, there were no indications that self-assessment was detrimental. In addition, previous studies have demonstrated benefits of self-assessment in the middle school chorus

classroom, such as improved student-to-teacher communication and improved student confidence (Matheny, 1994, McCall, 2007).

Self-assessment has customarily been used in academic classes and its use is now more widespread in arts education. Teaching music should be a reflective and ever-evolving practice. As studies demonstrate the benefits or drawbacks of more teaching tools, such as self-assessment and model-listening, it is the obligation of educators to apply this knowledge to their practice so that they can provide their students with the best music education possible.

REFERENCES

Anderson, J. (1981). Effects of tape-recorded aural models on sight-reading and performance skills. *Journal of Research in Music Education, 29*(1), 23-30.

Barefoot, R. (2006). Self-analysis skills for the developing singer. *Music Educators' Journal, 92*(3), 50-54.

Burrack, F. (2002). Enhanced assessment in instrumental programs. *Music Educators' Journal, 88*(6), 27-32.

Copland, A. (July, 2009). Ching a ring chaw [mp3]. Retrieved from http://www.arthursmiddle.com/music/choir7.html

Darrow, A. & Marsh K. (2006). Examining the validity of self-report: middle-level singers' ability to predict and assess their sight-singing skills. *International Journal of Music Education, 24*(1), 21-29.

Emerson, R. (1999). Shoshone love song. On *1999 All-state Chorus* [CD]. Dover, DE: AMP Recording & Duplicating Service.

Gardner, H. (1989). Zero-based arts education: an introduction to Arts PROPEL. *Studies in Art Education, 30*(2), 71-83.

Goolsby, T. W. (1999). Assessment in instrumental music. *Music Educators' Journal, 86*(2), 31-35, 50.

Goolsby, T. W. (1995). Portfolio assessment for better evaluation. *Music Educators' Journal, 82*(3), 39-44.

Gray, C. (2009). Song of the wind [choral octavo]. Dayton, Ohio: Lorenz Publishing Company.

Gray, C. (2009) Song of the wind. On *Song of the Wind Performance/Accompaniment* [CD]. Dayton, Ohio: Lorenz Publishing Company.

Hewitt, M. P. (2001). The effect of modeling, self-assessment, and self-listening on junior high instrumentalists' music performance and practice attitude. *Journal of Research in Music Education, 49*(4), 307-322.

Hewitt, M. P. (2002). Self-assessment tendencies on junior high instrumentalists. *Journal of Research in Music Education, 50*(3), 215-226.

Kizlik, B. (August 4, 2009). Measurement, assessment, and evaluation in education. Retrieved from http://www.adprima.com/measurement/html

Kruse, N. (2006). The effect of instruction on sixth grade band students' ability to self-rate etude performance. *Visions of Research in Music Education, 8*, 1-21.

Larkin, M. H. (1985). *The construction and validation of rating scales for the objective measurement of five dimensions of achievement in choral music* (unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Matheny, J. (1994). A grading system for secondary music groups. *Music Educators' Journal, 80*(4), 37-39.

McCall, M. S. (2007). Portfolio assessment in middle school chorus: students and teacher learning. Retrieved from ProQuest Digital Dissertations. (AAT 3249215).

McMillan J. H. & Hearn, J. (2009). Student self-assessment: the key to stronger student motivation and higher achievement. *The Education Digest, 74*(8), 39-44.

Molnar, J. W. (1948). Self-assessment by students. *Music Educators' Journal, 34*(4), 48, 52.

Morrison, S. J, Montemayor M., & Wiltshire, E. S. (2004). The effect of a recorded model on band students' performance self-assessments, achievement, and attitude. *Journal of Research in Music Education, 52*(2), 116-129.

Patterson, M. (July, 2009). This shall be for music [mp3]. Retrieved from http://www.arthursmiddle.com/music/choir7.html

Robinson, C. (1990). Choral performances: do you hear what I hear? *Music Educators Journal, 77*(4), 47-51.

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

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

Shaw, K. (July, 2009). Rock a my soul [mp3]. Retrieved from http://www.pjschools.org/district/NewsArchives/NewsArchives_08.09/district

Stamer, R. A. (2002). Choral ensembles for independent musicianship. *Music Educators Journal*, *88*(6), 46-53.

Swift, F. (1958). *A workbook in the fundamentals of singing.* Miami, Florida: Belwin Mills.

Thomas, A. (1999) Keep Your Lamps. On *1999 All-state Chorus* [CD]. Dover, DE: AMP Recording & Duplicating Service.

Wiggins, G & McTighe, J (2006). *Understanding by design.* Upper Saddle River, New Jersey: Pearson Education.

Appendix A

TEACHER CHORAL RATING SCALE

5 Point Scale
5=Superior
4=Excellent
3=Good
2=Fair
1=Poor

Adjudicator_____

Date_____

Group_____

Tone Quality

_____Ensemble sings with resonant Tone Quality
_____Vowel quality consistently is correct
_____Tone Quality is not nasal or breathy

Intonation

_____Pitches are performed accurately
_____The ensemble performs with an accurate sense of tonality
_____Breath management is adequate to maintain consistent Intonation

Rhythmic Accuracy

_____Entrances and releases are executed together
_____Ensemble performs accurately the melodic rhythm
_____Ensemble performs with a correct sense of meter

Expression

_____Text is clearly understood and performed with emotional understanding
_____Dynamics are performed accurately and stylistically
_____Ensemble performs in a tempo that is technically and stylistically appropriate

Balance and Blend

_____Melody is heard distinctly
_____Ensemble sings with homogeneity of Tone Quality within and among voice parts
_____The ensemble is not dominated by one or more voice parts

Total Score_____/75

Note: This rating scale is an adapted version of *The Five Dimensions of Achievement in Choral Music Performances* by Larkin, M. H. (1985).

Appendix B

STUDENT CHORAL RATING SCALE

5 Point Scale
5=Superior
4=Excellent
3=Good
2=Fair
1=Poor

Date_____

Group_____

Tone Quality: nice sound, not nasal or breathy, pure vowels

1 2 3 4 5

Comments:

Intonation: accurate pitches

1 2 3 4 5

Comments:

Rhythmic Accuracy: entrances and releases together, correct sense of meter and tempo

1 2 3 4 5

Comments:

Expression: clear diction, text meaning portrayed in music, expressive performance

1 2 3 4 5

Comments:

Balance and Blend: melody heard distinctly, individual voices and voice parts do not stick out

1 2 3 4 5

Comments:

Total Score_____/25

Note: This rating scale is an adapted version of *The Five Dimensions of Achievement in Choral Music Performances* by Larkin, M. H. (1985)

Appendix C

September 14, 2009

Dear Parents,

As part of obtaining my Master's degree in music education at the University of Delaware, I am conducting a study on the effects of model-listening and student self-assessment on middle school chorus students' achievement. I will be conducting this 10-week study at Newark Charter School during the Fall 2009 semester with the Concert Choir. For the study, the Concert Choir members will be randomly assigned to four "mini-choirs," each containing a balance of voice parts (soprano, alto, etc.). Group one will periodically listen to recordings of a model choir; group two will engage in self-assessment; group three will engage in both model-listening and self-assessment; and group four will be doing neither. During the 10th week, each of the four ensembles will be assessed to see if there is a correlation between model-listening and/or self-assessment with choral performance achievement.

This study will involve the making of audio recordings so that the self-assessment groups may listen to and assess their performances. Also, each ensemble will be recorded at the end of the study so that two Newark Charter music teachers (Angela Sheik and Sarah Aherne) and I may evaluate and compare their progress. I am writing to make sure you do not object to having your child's singing as part of a small ensemble (of about 15 students) be audio-recorded. The audio recordings will be kept confidential and they will be erased after the results for this study are obtained.

If you **DO NOT** want your child participating in this study, please sign and return the bottom portion of this letter to me by **Monday, Sept 21st**. Please note that this study will take place only during regularly scheduled chorus rehearsals and that no additional time or effort on a part of the students will be required or necessary. Feel free to contact me with any questions.

Thank you,
Jenny Ward
Music Teacher

~~~~~~

I do not want _____ to participate in this study.
(Child's Name)

Signed,

_____ _____
(Parent or Guardian) (Date)
~~~~~~

Appendix D

CITI Collaborative Institutional Training Initiative

Course In The Protection Human Subjects Curriculum Completion Report
Printed on
Learner: Jenny Ward (username: jward417)
Institution: University of Delaware
Contact Information 513 Paisley Pl
Newark, DE 19711 USA
Department: Music education
Phone: 302 369-2001
Email: jward417@yahoo.com

Graduate Students:
Stage . Basic SBR Passed on 06/12/09 (Ref # 2892175)

Required Modules	Date Completed	Score
Belmont Report and CITI Course Introduction	06/12/09	3/3 (100%)
Students in Research – SBR	06/12/09	10/10 (100%)
History and Ethical Principles – SBR	06/12/09	4/4 (100%)
Defining Research with Human Subjects – SBR	06/12/09	5/5 (100%)
The Regulations and The Social and Behavioral Sciences – SBR	06/12/09	5/5 (100%)
Assessing Risk in Social and Behavioral Sciences – SBR	06/12/09	5/5 (100%)
Informed Consent – SBR	06/12/09	4/4 (100%)
Privacy and Confidentiality – SBR	06/12/09	4/4 (100%)
Research in Public Elementary and Secondary Schools – SBR	06/12/09	4/4 (100%)
Conflicts of Interest in Research Involving Human Subjects	06/12/09	2/2 (100%)
University of Delaware	06/12/09	no quiz

Elective Modules	Date Completed	Score
Research with Children – SBR	06/12/09	25

For this Completion Report to be valid, the learner listed above must be affiliated with a CITI participating institution. Falsified information and unauthorized use of the CITI course site is unethical, and may be considered scientific misconduct by your institution.
Paul Braunschweiger Ph.D.
Professor, University of Miami
Director Office of Research Education
CITI Course Coordinator

Appendix E

RESEARCH OFFICE
210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: August 6, 2009

TO: Jenny Ward, BM

FROM: University of Delaware IRB

STUDY TITLE: The Effects of Self-Assessment and Model-Listening on Middle School

Chorus Students' Achievement

IRB REFERENCE #: 128743-1

SUBMISSION TYPE: New Project

ACTION: DETERMINATION OF EXEMPT STATUS

DECISION DATE: August 6, 2009

REVIEW CATEGORY: Exemption category # 1

Thank you for your submission of New Project materials for this research study. The University of Delaware IRB has determined this project EXEMPT FROM IRB REVIEW according to federal regulations. We will put a copy of this correspondence on file in our office. Please remember to notify us if you make any substantial changes to the project. If you have any questions, please contact Elizabeth Peloso at 302 831-8619 or epeloso@udel.edu. Please include your study title and reference number in all correspondence with this office.