NORTHWESTERN UNIVERSITY

The Effects of Feedback on Sight-singing Achievement

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Music

By

Adam G. White

EVANSTON, ILLINOIS

December 2020

Abstract

Sight-singing skills are a key component in empowering choristers to make music independently. Advances in technology have made visual feedback displaying the accuracy of a sight-singing attempt available to singers. The purpose of this study was to examine the effects of computerized visual feedback provided by the *SmartMusic* interface on sight-singing achievement among choristers (n = 77) from two suburban high schools. Over a period of five-weeks, choristers engaged in weekly sight-singing assessment sessions where they sight-sang a melody, reviewed that melody for 90 seconds, then sang that melody again. Using a matched-group design, participants were assigned one of three groups: those viewed feedback following their initial attempt, those who viewed feedback following their follow-up attempt, and those who did not view any feedback. Sight-singing scores were evaluated by group for improvement during each assessment session, and from pretest to posttest. Results determined that while students made significant improvements on a melody following a sight-singing attempt, those improvements were not affected by feedback condition. Posttest scores were not significantly higher than pretest scores for any group. These findings suggest that though feedback may be an important component in the development of sight-singing skills, the computerized feedback provided in this study was no more effective than receiving no feedback at improving sight-singing achievement. Findings also suggest that teachers should use this available feedback to adjust the difficulty of sight-singing assessments to fit the ability levels of students. Furthermore, students were unable to transfer learning from practice with a click-track and note indicator to performance without these features so teachers should design summative assessments to match

the task presented during formative assessments. This technology may be best utilized as a supplement to good teaching but is not designed to replace a quality teacher.

Acknowledgements

There are so many who have helped me along this journey. Thought I cannot address them all, I would like to take a moment to express my gratitude. First and foremost, I am grateful for Barbra, Henry, and Norman who were willing and eager to uproot our lives and begin a new adventure in Evanston. I love you and you earned this degree right along with me.

Thank you, Sarah Bartolome, for seeing us through the most difficult times with caring and compassion. You held the department together when Maud retired and Steve passed away. I will be forever grateful (as will future generations of doc students!) because you showed us how to treat doctoral students with dignity and respect.

To my committee: Sarah Bartolome, Steven Demorest, Maud Hickey, Steven Morrison, and Richard Ashley, thank you for sharing your vast knowledge with me throughout this project. I will be using the skills you taught me for years to come! Thank you, Kansas choral directors, Jon D., and Cassie B., Dr. Dwayne Dunn, and Randy Frye who made this study possible. A very special thanks to Aaron Himes for providing expert assistance when I needed it most. Thank you, Jeremy Manternach, James Daugherty, Ed Raines, and Martin Bergee for helping me get into doctor college. Thank you, Anita Brookens, Susie Lewis, Veda Journagan, and Jon Young for spending so much time in my classroom. What we accomplished, we did together.

Amanda Draper, Casey Schmidt, Diego Pinto, and Patrick Horton, I love you all. Thanks for always being there for me. X^2 forever. Let's get together at a conference (remember those?)! Thank you, Brian Weidner, Jake Berglin, Erica Sagerman, Seth Adams, David DeAngelis, Miguel Garcia, and Aaron Himes for being colleagues and speaking your mind on Thursday afternoons and ACF evenings.

Dedication

For Steve Demorest

"If something is worth doing, it is worth doing well.

If you can't do it well, that's okay.

Get better."

and

I should have run a Chi-square.

Table of Contents

Chapter 1

Introduction

A primary goal of any educational endeavor is to develop independent learners capable of solving problems. In music, the ability to read, perform, and interpret music written in Western notation is a critical skill for individuals in many ensembles. The ability to sing a written melody at sight without the help of an external pitch reference is an elusive task for novice and experienced musicians alike, but once achieved, that ability represents a high level of musical independence. Effective sight singers can open a score, unlock the music contained within, and access the wonderfully broad diversity of written music. Since it is such a valuable skill, music educators should continue to teach sight-singing as researchers continue to seek new ways to enable them to do that job well.

Notational music literacy, which includes sight-reading and sight-singing, involves both musical independence and problem-solving skills. Scholars have defined sight-singing, and the tasks involved with it, in a multitude of ways. In his book Comprehensive Choral Music Education, John Hylton (1994) described sight-singing as an individual's ability to take a piece of choral music and accurately produce the pitches, rhythms, and expressive markings without the aid of a piano or external instrument (1994). Additionally, Vujović and Bogunović (2012) describe sight-singing as,

> An "online" activity that asks for quick insight and problem solving in order to maintain fluency and accuracy. The sequence of events is: perceiving notation, processing it, and executing the resulting motor (or vocal) program. During sight-reading-singing, pattern

> recognition is in progress and it is related to long term working memory which supposes that experts are able to access quickly the contents of their long-term memory. (p. 1106)

W. G. McNaught (1899) summarized some late Nineteenth Century beliefs about sight-singing. He concluded that there are three main areas of concentration: (a) the memory of pitch, either permanent (absolute pitch) or temporary, (b) interval effects, and (c) the sense of key. Raymond Mosher (1925) found seven factors that contribute to an individual's ability to sight-sing: (a) an understanding of music notation and symbols, (b) the ability to recognize scales, chords, and intervals, (c) the ability to understand rhythmic value, (d) the ability to name well-known melodies at sight, (e) the ability to complete harmonic dictation, (f) the ability to complete rhythmic dictation, and (g) the ability to complete melodic dictation. Rose Dwiggins (1984) focused on the following concepts related to sight-singing: discrimination of notation and knowledge of music signs, chord analysis, and pitch discrimination. James Middleton (1984) believed that sight-singing could be separated into two distinct sets of vocabularies, rhythmic and tonal. Sight-singing is a complex task including visual, motor, and aural skills. The present study seeks to contribute to this centuries-old area of inquiry, examining the impact of feedback on sight-singing achievement of high school aged choristers. For the purposes of this study, sight-singing will be defined as the accurate reproduction of pitches and rhythms with the voice from traditional Western written notation.

History of Sight-singing Instruction

As long as singers have been performing music from notation, sight-singing has been an important skill. According to Allan Atlas (1998), Medieval music theorist Guido d'Arezzo (c. 991-after 1033) found young singers were able to learn new melodies at a higher rate if they sang a group of syllables that related to certain pitches. These syllables, *ut, re, me, fa, sol, la,* became the first solmization system on record. Over 600 years later, a four-syllable solmization system was common in England and in the *Bay Psalm Book* (1651) of Puritan New England. This method used *fa, so, la,* and *mi*, and was aptly named *fasola*. This method remained popular in the United States until the mid-nineteenth century (Keene, 2009).

Henrich Pestalozzi (1746-1827) was a Swiss education reformer who set out to "elevate the lowly condition of the common people" (Mark, 1994, p. 91). His philosophy of "learning by head, hand, and heart" had a profound effect on the creation of new schools in the United States. By 1834, the Bostonian singing instructor Lowell Mason had written the book, *Manual of the Boston Academy of Music, for Instruction in the Elements of Vocal Music, on the System of Pestalozzi*. With this text, Mason outlined the use of a seven-syllable solfege method and advocated the use of rote teaching prior to instruction in music literacy. The sound before sight concept was also paramount to Hungarian composer, musician, and educator, Zoltán Kodály (Ittzés, 2004).

The Kodály Method emphasized folk songs and utilized solfege syllables to aid in the familiarity of pitch relationships. This method also encouraged the use of hand-signs representing solfege syllables (Demorest, 2001).

School choirs of the mid-nineteenth century in the United States were instituted primarily as a means of teaching sight-singing (Mark & Gary, 1992). By the early twentieth century, these early music literacy courses were replaced by glee clubs and other choruses often comprised of the entire student body. By the 1920s, much of the previous emphasis on singing was replaced by interest in band and orchestra (Kegerreis, 1970). More recent trends in music education have indicated a resurgence of interest in music literacy as evidenced by the mass of published materials on sight-singing, the increase of sight-singing in contests and festivals, and an emphasis on standards and assessment (Demorest, 2001).

The National Association for Music Education (NAfME), formerly known as the Music Educator's National Conference (MENC), identified reading and notating music one of their nine standards. Standard 5b stated, "Students who participate in a choral or instrumental ensemble or class sight-read, accurately and expressively, music with a level of difficulty of 3, on a scale of 1 to 6" (1994). The current Core Arts Standards, adopted by NAfME in 2014, include MU:Pr4.2.E.5a which states, "Demonstrate, using music reading skills where appropriate, how knowledge of formal aspects in musical works inform prepared or improvised performances." The Repertoire and Resources Committee for Children and Community Youth Choirs, part of the American Choral Directors Association, has published a set of seven standards. One of these standards is titled, "Rehearsal techniques and Instruction," with the secondary heading "Literacy." The first enumerated bullet states:

> Ensembles should be given the knowledge to be independent musicians. Literacy is important to independence and should be taught in the manner that is most comfortable for the conductor, whether that be through solfege, available sight-singing resources (books), or through the repertoire (ACDA, retrieved 4/14/2019).

In addition, adjudicated sight-singing has also been a part of festivals and honor ensemble auditions in the United States. Norris (2004) found that 24 states (48%) included a formal sight-singing requirement in state-level high school choral ensemble adjudications. Several studies (Demorest, 2001; Snider, 2007; Brendell, 1996) have shown that the existence of sight-singing at festivals tended to have a positive correlation with time teachers spent on sight-singing instruction. According to Matthew Armstrong (2001) there exists further incentive for a choir director to teach sight-singing: “By including choral sight-reading as a part of the adjudications process, we afford ourselves as choral directors an incentive to remain accountable, and provide ourselves and our students an opportunity to benefit from the expertise of other professional musicians” (p. 29).

Chorus teachers value sight-singing skill development among their choristers and feel sight-sing instruction is an important component in the rehearsal process, though more directors believe in the efficacy of sight-singing instruction than actually teach it (Von Kampen, 2003; Myers, 2008; Farenga, 2013; Potter 2015). Some choir directors believe the process of learning sight-reading skills improves their choir’s overall intonation (Floyd & Bradley, 2006). Schools of music have used sight-singing as a component for program admissions and proficiency as a requirement for graduation (Hime et al., 2014), though it is unknown how many schools have sight-singing standards and how stringent these standards may be.

The long history of sight-singing instruction, its appearance in national standards and policy, the existence of adjudicated sight-singing, and documented chorus teachers’ beliefs, all demonstrate the notion that sight-singing is a valued skill in vocal music education. The value of sight-singing historically and in music education policy and practice suggests that continued investigation into best practices surrounding sight-singing instruction would be beneficial to the field, especially within the ever-evolving context of contemporary choral music education.

Teaching Sight-singing Effectively

Empowering students with music literacy skills is a positive step toward building chorister independence and agency for learning music literature. Middleton (1984) believed that choir directors and performers alike find instruction in music literacy beneficial: “Directors can provide the most beneficial choral experiences to student by giving them access to music knowledge, concepts, and skills – not the least of which are music-reading skills” (p. 31).

To encourage sight-singing skills among choristers, a choir director should develop a plan to teach the necessary skills for sight-singing. Researchers have studied the effectiveness of different strategies for teaching and learning sight-singing, including pedagogy (Benton, 2002; Boisen, 1982; Killian, 1991; Kostka, 2000), systems (Brown, 2001; Demorest & May, 1995; Henry & Demorest, 1994; McClung, 2008), and the need for individual assessment (Demorest, 1998; Nolker, 2006). Though research on the use of various methods and solmization systems is mixed, individual assessment has been found to be an effective way to improve chorister sight-singing achievement.

Sight-singing Assessment

In order to ensure individual learning has taken place, chorus teachers need to assess their students individually to understand what they know. Peggy Bennett (1983) argued that class instruction and practice may not teach individuals to sight sing. She found students were quick to imitate the pitches they hear, often within a fraction of a second. This imitation could lead an experienced choir director to believe that the entire choir can accurately sight-sing, when it may be as few as one or two students leading the group. Research literature has also found individual sight-singing assessment improved student performance (Demorest, 1998) and ensemble success does not indicate individual sight-singing ability is present (Nolker, 2006). Therefore, assessing individual student sight-singing abilities is necessary to encourage all choristers to learn those skills.

Traditional methods for assessing individual sight-singing have often been cumbersome and time consuming for chorus directors. Recorded assessment, quartet testing, and early computer testing required rehearsal time and teacher-produced feedback. As Steven Demorest (2001) suggested, “The key is to find an assessment procedure that makes sense and use it consistently” (p.123). However, directors have reported a perceived lack of time to complete individual assessment, despite knowing its efficacy (Goss, 2010). It is possible that advances in technology have offered choir teachers more efficient alternatives to traditional methods of assessment as evidenced by the increase in the reported use of technology for individual assessment (Hawkins, 2018; Henry, 2015).

During the early stages of the COVID-19 pandemic, teachers were scrambling to find ways of meeting their student learning outcomes remotely with the assistance of technology. Some companies who marketed music learning software including *MakeMusic* (2020) and *Sight Reading Factory* (2020), offered their products for free for teachers who were looking to adapt to an online learning environment. It is unclear at this time how many teachers made use of these products, but an interest utilizing technology as a supplement to instruction became evident (Chrysostomo & Triantafyllaki, 2020).

Feedback

A missing piece from our current knowledge base about sight-singing instruction is the efficacy of feedback during individual assessment. One of the unique qualities of available technology is an assessment feature that provides visual feedback on Western music notation. It has been suggested that "individuals acquire a skill much more rapidly if they receive feedback about the correctness of what they have done. If incorrect, they need to know the nature of their mistake" (Pellegrino et al., 2003, p. 85). It is pertinent, then, to study feedback as a dependent variable in sight-singing research. The current affordability and availability of computer programs like *SmartMusic* that offer a feedback feature make that research timely.

Mory (2004) discussed five types of feedback variables that may be of interest in feedback research: (a) *No feedback* (the absence of feedback); (b) *Simple verification feedback* (indicates correct and incorrect responses); (c) *Correct response feedback* (informs what the correct response should have been); (d) *Elaborated feedback* (provides the cause of the error); and (e) *Try-again feedback* (allows for an additional attempt). This study manipulated the presence of *SmartMusic* feedback, which provides *simple verification feedback* and *correct response feedback,* by changing the timing of that feedback or by presenting *no feedback*.

In conclusion, music educators and choir directors have historically valued sight-singing skills that can empower choristers to make music independently, allowing for access to a broad diversity of repertoire written in Western notation. Various sight-singing methods, strategies, and solmization systems have been studied and most do not reveal significant differences in sight-singing achievement. Among those strategies studied, however, individual assessment has been found to be an effective tool in developing sight-singing skills, but we do not know which among its component parts contribute to that effectiveness. This study provides insight into the assessment process by examining feedback, a small but potentially important piece in learning sight-singing skills.

The following chapters will delineate the details of this study. In Chapter Two, I address the current body of sight-singing research and provide a detailed rationale for this study. Chapter Three contains a detailed description of the methods, materials, and data collection I used while Chapter Four presents the findings from my statistical analyses. In Chapter Five I discuss the findings, exploring and interpreting the results, making connections to the extant literature, forwarding implications for choral music educators, and suggesting directions for future research.

Chapter 2

Literature Review

Sight-singing has been a topic of study by music education researchers since the first volume of the *Journal of Research in Music Education* in which Hutton (1953) published a study in which audio-visual materials significantly increased the effectiveness of sight-singing instruction. In order to situate the present investigation in the landscape of existing research relating to sight-singing, this literature review will synthesize and interpret studies relating to (a) predictors of sight-singing achievement, (b) sight-singing instructional practices, (c) sight-singing cognitive processes, (d) eye-gaze during sight-singing tasks, (e) sight-singing pedagogy, (f) individual assessment of sight-singing, and (g) technology use during individual sight-singing assessment. I will also include a brief overview of the use of feedback in education research.

Predictors of Sight-singing Achievement

Various studies found particular factors that were predictive of sight-singing success. Among these factors were musical background, musical aptitude, aural skills, and academic achievement. These studies are important for understanding and explaining differences in sight-singing abilities.

Colwell (1963) found that instrumental students scored, on average, higher than vocal students on a sight-singing task. Students with a background in piano evidenced the highest level of sight-singing achievement. Tucker (1969) found the following experiences and combinations of experiences were predictive of sight-singing success. Listed in order from high achievement to low achievement, were: (a) students with at least six years of piano lessons plus vocal and instrumental experience, (b) students with at least six years of piano lessons plus instrumental experience, (c) students with at least six years of piano lessons plus vocal experience, (d) students with instrumental experience, (e) students with vocal experience, (f) students with general music experience only, and (g) students with no musical experience. Demorest and May (1995) found strong correlations between sight-singing success and the number of years a student was in choir, the number of years playing piano, and private music instruction. Furby (2008) investigated the sight-singing backgrounds of incoming undergraduate students auditioning for a university choral ensemble. Participant background variables were regressed on sight-singing achievement. Years of high school choir participation was the greatest predictor of sight-singing success though students who reported having undergone sight-singing training scored significantly higher on their audition than those who did not report such training.

Harrison (1990) studied correlations between musical aptitude, academic ability, music experience, and aural skills assessment scores in music students' sight-singing abilities in a freshman music theory course. The strongest positive correlations predicting successful sight-singing were academic ability and music experience. Harrison et al. (1994) studied the relationships between (a) musical aptitude, (b) academic ability, (c) music experience, and (d) motivation and the aural skills of university music theory students. Musical aptitude, academic ability, and music experience were predictors of success in ear-training and sight-singing skill development while student motivation did not appear to be predictive of these skills.

Relationships were found between the sight-singing abilities and melodic dictation skills of freshman music majors (Thostenson, 1967). Norris (2003) also investigated possible correlations between sight-singing ability of freshman university students and the ability to complete melodic dictation accurately. Participants completed a pretest, and following a semester of an aural skills class, completed a posttest. Melodic dictation was found to be a moderately strong predictor of sight-singing achievement. Larson (1977) examined possible correlations between three factors: (a) melodic error detection, (b) sight-singing, and (c) melodic dictation among junior and senior music majors. Participants were given diatonic, chromatic, and atonal passages to sight-sing. A stronger positive relationship was found between error detection scores and melodic dictation scores than between error detection and sight-singing scores.

In summary, the extant literature suggests that instrumental experience is a strong predictor of sight-singing success though continued choir participation is also a factor. Various aural skills, including melodic dictation and error detection, also predict success. These factors have contributed to a wide variety of sight-singing abilities among choristers in secondary ensembles (Daniels, 1986).

Descriptive Studies of Sight-singing Instructional Practices

Researchers have primarily relied on self-report data to uncover sight-singing practices among primary, secondary, and university chorus instructors. Research on instructional practices has provided a better understanding of what is happening in music classrooms.

Time spent on Sight-singing Instruction

A web-based survey indicated that responding secondary teachers spent an average 9.5 minutes per rehearsal on sight-singing instruction (Demorest, 2004). Among university choir directors, 64.5% reported teaching sight-singing to their ensembles, while 93.4% believed sight-singing instruction should be part of rehearsals (Myers, 2008). Secondary choir directors have reported spending between 18% (Floyd & Bradley, 2006) and 22.2% (Brendell, 1996) of rehearsal time on sight-singing instruction. A sample of Arizona choral teachers with 20 or more years of experience reported significantly more agreement with the statement, "Sight-singing is a part of my choir rehearsal." Arkansas middle and high school directors reported spending 15% of their rehearsal time on sight-singing instruction (Fisher et al., 2015).

Use of Solmization and Rhythm Systems

Solmization systems, as a method for teaching pitch relationships have been found to be an important component in the development of sight-singing skills. Teachers have reported using a variety of these systems and the results vary by grade level and geographical region. Demorest (2004) found the preferred method (64%) of pitch reading was movable *do.* Other methods included the use of numbers (21%) and the use of fixed *do* (15%) (Demorest, 2004). Middle school teachers reported using movable *do* without hand signs (40.0%) more frequently than solfege (fixed or movable *do*) with hand signs (33.6%), numbers (14.9%), fixed *do* (6.9%), and neutral syllables or letter names (4.6%) (Nichols, 2012).

Floyd and Bradley (2006) also found movable *do* to be the preferred method by 75% of participants as did McClung (2001), who found movable *do* was the most prominent system used in choral rehearsals across a six-state region among high school all-state choristers. High school choristers from Texas reported movable *do* (80%) was used more frequently than fixed *do* (12.1%), and numbers were reported infrequently (Henry, 2013). Among middle school teachers (Nichols, 2012), movable *do* without hand signs (40%) was reported more frequently than solfege (fixed or movable *do*) with hand signs (33.6%), numbers (14.9%), fixed *do* (6.9%) and neutral syllables or letter names (4.6%).

The prevalence of rhythm-reading systems has garnered less attention from researchers. Middle school choir teachers reported using a broad diversity of systems (Nichols, 2012) while

Demorest (2004) found counting (47%) to be most common system among respondents while the remainder (53%) reported using other methods including neutral syllables, *ta-ti-ta*, or *Takadimi*. Arkansas teachers preferred counting (67.2%), rhythm syllables (27%), and other methods (14%) for teaching rhythms (Fisher et al., 2015).

While self-report surveys offer a broad snapshot of instructional practices, they can be problematic. Sight-singing research was often influenced by selection bias and choir teachers tended to overstate the amount of time spent on sight-singing instruction (Demorest, 2001). To date, only Brendell (1996) relied on recorded observations of sight-singing instructional practices. Moveable *do* was the most common solmization system reported in the United States, and it appears as though more teachers were choosing movable *do* more recently (Henry, 2014), than were ten years earlier (Demorest, 2004). Rhythm systems were more diverse and did not have the same depth of study as solmization systems.

Sight-singing Cognitive Strategies

Numerous studies have investigated cognitive processes used by those engaged in sight-singing activities to better understand the task's complexities. A qualitative investigation sought to understand the cognitive strategies used while sight-singing (Fournier et al., 2017). Researchers examined extant literature, textbooks, sight-singing manuals, and interviewed music students and aural skills teachers revealing four broad categories: (a) reading mechanisms, including pitch decoding, pattern building and validation; (b) sight-singing, including preparation and performance; (c) reading skills acquisition, including vocabulary development, symbolic association, internalization, and rehearsal techniques; and (d) learning support, including selfregulation, time-management, attention, stress, and motivation. This inventory provided a framework for understanding the components of sight-singing that may be taught or researched.

An examination of participant meta-cognition while sight-singing sought to determine which cognitive strategies were used (Vujocić & Bogunović, 2012). Results showed sight-singers assessed easy and difficult passages then defined how they would go about performing more difficult passages. Participants with higher scores believed they were effective sight singers, relied on audiation, retained tonic, and analyzed the score more before singing. Knox (2003) compared mistakes or miscues made by vocalists while sight-singing and reading text and determined that sight-singing utilized the same mental processes and cognitive strategies as reading, and that both sight-singing and reading text formed a semiotic system.

Sight-singer's Eye Gaze

An examination of singers' gaze while sight-singing determined that skilled sight-singers looked farther ahead in the notation and returned to the point of performance in the music (Jacobssen, 1942). Less-skilled sight singers' eyes tended to fixate for longer periods of time and generally remained at the point of performance (Goolsby, 1994a, 1994b). Speed and accuracy of sight-singing was related to the number and duration of fixation pauses and the number of regressions. In rehearsal settings, eye-tracking apparatus worn by choral singers reading an unfamiliar piece revealed that novice singers had difficulty keeping gaze on their melody line and tended to fixate longer on large intervals of a 4^{th} and 5^{th} (Timoshenko, 2018). As a whole, these studies help our understanding of the nature of sight-singing processes and assist in designing our instruction.

Sight-singing Pedagogy

This section will review research related to sight-singing pedagogy, excluding assessment research which will be addressed in a later section. Researchers have investigated the efficacy of specific sight-singing solmization systems, practices, methods, and strategies, to provide empirically derived information to inform best practices.

Solmization Systems and Hand-signs

Researchers have studied various solmization systems in search of which ones are the most effective at teaching sight-singing. Brown (2001) studied the effectiveness of movable *do* and fixed *do* in the sight-singing of diatonic, modulatory, chromatic, and atonal melodies among undergraduates enrolled in a music theory course. Results suggested that the movable *do* group was significantly more effective on the performance of chromatic music. While the fixed *do* group was significantly more effective on the performance of atonal passages and passages with a high difficulty level, the overall difference in sight-singing achievement between the two groups was not significant. Henry and Demorest (1994) also compared individual sight-singing abilities among high school students from two schools. One school used a fixed *do* system, and the other used a movable *do* system. Both choirs had received outstanding sight-singing ratings at a state contest for at least three years, and there were no significant differences between the sight-singing abilities of students from the two schools. Demorest and May (1995) compared the sight-singing success of individual students from four Texas high schools using fixed *do* and movable *do* sight-singing systems. Differences between the two systems were attributed to factors other than the system used. A group of students who wrote solfege syllables in the score was compared to another group that used the time for audiation (Lovorn, 2016). Both groups scored significantly higher on a posttest but differences between groups were not significant.

Cassidy (1993) compared the sight-singing abilities of five different treatment groups using echo singing combined with (a) solfege with hand-signs, (b) solfege alone, (c) letters of the notes, (d) "la" as a neutral syllable, and (e) nothing to serve as a control group. Comparisons of pretest and posttest data yielded significant improvement within the solfege only group and the solfege with hand signs group, suggesting the effectiveness of solfege, and solfege when accompanied with hand signs. Other research on hand-signs' efficacy has found they did not improve sight-singing achievement (Frey-Clark, 2017). McClung (2008) conducted the most recent research to investigate the effects of hand-signs on high school choristers' sight-singing abilities. Among participants who had extensive training in the use of hand signs, no significant differences were found between a group of students who used hand-signs while sight-singing and a group that did not use hand-signs while sight-singing. McClung concluded that hand-signs might not be effective for all, but some singers may benefit from voluntary use.

Other Strategies

While solmization systems have garnered a good deal of focused attention from researchers, other studies have examined a variety of practices and how they impact sight-singing achievement. Benton (2002) compared the effects of metacognition on middle school students' sight-singing achievement and attitudes toward sight-singing instruction. The treatment groups were given activities in the following areas: (a) think-aloud activities with partners, (b) self-assessment activities, and (c) self-reflection activities. Among the seventh-grade participants, the control group scored significantly higher on a posttest. However, the treatment group demonstrated more positive attitudes about sight-singing. Among the eighth-grade participants, neither group scored significantly higher on a sight-singing posttest. The eighth-grade treatment group also demonstrated more positive attitudes about sight-singing instruction.

A comparison was made of sight-singing abilities among elementary education majors who received piano instruction alone and those who received sight-singing training as part of piano instruction (Hargiss, 1962). Results showed that students who had sight-singing training combined with piano instruction scored significantly higher on a sight-singing performance test than those who had received piano training alone. Lucas (1994) examined differences in sight-singing achievement related to harmonic context, either (a) melody only, (b) piano harmony, or (c) vocal harmony. Results found a melody-only context was the most effective way to practice sight-singing among middle school choral students.

In a 2004 study, Henry compared two groups of inexperienced high school singers to determine the effects of learning intervals and pitch relationships (targeted pitch skills) using familiar and unfamiliar melodies on sight-singing achievement. When using familiar melodies to teach pitch skills, students learned intervals by relating them to popular songs. Results indicated significant improvement for both groups but no differences between groups. Henry (2011) later examined targeted pitch and rhythm skills to study the interaction between difficult pitch passages and difficult rhythm. Results suggested difficult rhythm did not affect ability to sing the correct pitch, but difficult pitch affected rhythm. Rhythm success was a predictor of pitch success, suggesting that pitch often followed suit when rhythm was correct. In general, singers were more successful with pitch skills than rhythm indicating singers prioritized pitch over rhythm. Henry (2013) also studied the effect of key on sight-singing achievement. No significant differences were found among sight-singing attempts in D major, E-flat major, and F major.

In most sight-singing pedagogy research, no single practice, method, or solmization system is more effective than others. However, using a system is more effective than not using one. Additionally, engaging in sight-singing activities is more effective than not doing so (Demorest, 2004).

Sight-singing Assessment

Individual sight-singing assessment has been found to be an effective way to improve sight-singing achievement. Demorest (1998) conducted a study to determine if students given regular individual testing would score higher on sight-singing assessments than those given class instruction alone. Individual testing was found to be an effective way to increase scores in a classroom setting. Nolker (2006) investigated students' individual sight-singing abilities from choirs who received high sight-singing ratings and low sight-singing ratings. The results indicated that successful sight-singing ensembles had as many poor sight-singing students as unsuccessful sight-singing ensembles, concluding that group success was not an indicator of individual sight-singing ability.

Most directors (83%) reported they included some sort of assessment in their teaching (Demorest, 2004). Of those teachers who reported regular sight-singing assessments, almost half (47%) preferred formal to informal assessments. Preferred procedures for assessment were students performed alone, students performed in quartets, and students performed in a combination of solos and quartets. Goss (2010) surveyed middle and high school choir teachers to determine their self-described assessment practices. Directors reported an understanding of individual sight-singing assessment's efficacy, but found it difficult due to time constraints. Goss observed a need for research to find efficient sight-singing assessment procedures.

Killian and Henry (2005) compared the differences in individual sight-singing assessment scores between high school students who were given a 30-second practice before they sang and those who were not given that opportunity. Students with low accuracy scores did not benefit from preparation time, while those with higher accuracy did. The traits common in high scoring students were as follows: (a) they established the key, (b) they used hand signs, (c) they sang out loud during practice, (d) they finished the melody early and worked on problem areas, (e) they kept the beat with their bodies, and (f) they kept the beat steady. Traits found among ineffective students included: (a) lack of a steady beat, (b) stopping the melody, (c) taking eyes off the music, and (d) shifting their body position. Henry (2008) created a study were high school students were explicitly taught those behaviors. Participants were pretested and separated into two groups, low and high achieving. Both groups were instructed in desirable sight-singing behaviors. Low achieving students showed significant sight-singing improvement while gains in the high scoring group did not. Both groups demonstrated significantly more of the desirable behaviors following the treatment period. These findings suggested that students, especially inexperienced sight-singers, should be taught to practice desirable behaviors during sight-singing instruction to increase the likelihood they will use them.

Assessing sight-singing abilities is important to ensure students have demonstrated learning, and it is effective in improving student scores, but teachers have reported a lack of time to assess their students individually. Research has yet to discover which aspect of individual assessment, whether feedback, knowledge of results, or motivation, contributes to its effectiveness.

Technology-assisted Sight-singing Assessment

Researchers have been interested in the efficacy of technology as an assessment tool for several decades. Lorek (1991) compared the results of a computer-based vocal analysis software to a panel of sight-singing instructors. The vocal analysis software was found to be consistent with the instructors' judgments. Ozeas (1991) examined the effects of the computer program *Perceive* on university students' abilities to identify intervals, sing intervals, and sight at sight Results indicated that students given traditional instruction produced significantly greater sight-singing improvements than the computer training only group. Platte (1981) studied the effects of the computer program *Melodious Dictator* on university choral students' sight-singing abilities. Results indicated no significant differences in sight-singing abilities among students receiving computer training and those who did not.

Advances in technology have led to increased use of computerized assessment in the choral classroom (Hawkins, 2018). Among respondents, 66% used some form of individual sight-singing assessment, of which 46% reported using technology. Participants (19%) reported using computer-based performance evaluations, like *SmartMusic*, an increase from what was reported by Henry in 2015. Nielsen (2013) surveyed music teachers to determine their self-described use and attitudes of computerized assessments and found music teachers reported using technology-assisted computerized assessments for performance assessments, (a) *regularly* (8%), (b) *sometimes* (20%), (c) *occasionally* (42%), and (d) *never* (23%).

SmartMusic proprietary technology, including smart accompaniment and assessment features, has been of interest to music education researchers. It has been the focus of several instrumental music studies, investigating such topics as motivation (Gurly, 2012; Perry, 2014), attitudes toward use (Owen, 2015; Walls et al., 2013), assessment (Buck, 2008; Karas, 2005), and achievement (Flanigan, 2008). Among sight-singing assessment research, Henry (2015) examined high school students' perceptions of the efficacy of *SmartMusic* as a sight-singing assessment tool. All participants reported sight-singing was taught in their choral classroom with most (74.0%) reporting practicing *frequently*. Fewer students (9.4%) reported frequently practicing sight-singing on their own. Most students (70.2%) reported experiencing individual sight-singing assessments as part of their choral classroom, but computerized assessment experience was uncommon among participants (7.2%). Participants reported the highest anxiety when being assessed in front of peers, followed by assessment by computer. Assessments by recording and alone with a teacher were reported as less anxiety-inducing options. Results indicated a significant change in students' opinions about computerized assessment following the assessment treatment from *favorable* or *neutral* (68.8%) to *unfavorable* (68.8%), yielding a significant change $x^2(2) = 52.39$. The most common complaint reported was the inability to adjust tempo. Students who participated in the additional sight-reading class reported more favorable opinions of the *SmartMusic* assessment feature than those who did not. Henry concluded that the interface should be introduced to gain familiarity prior to starting sight-singing assessments using *SmartMusic*.

Petty and Henry (2014) compared the use of *SmartMusic* for sight-singing assessment feature to traditional assessment methods. Sixth grade beginning choir students were divided into a sight-singing technology group and a sight-singing without technology group. Using an eight-week treatment, students underwent individual sight-singing assessments using *SmartMusic* or paper notation. Both groups showed significant improvement, but no differences were found among groups. The researchers stated, “While it was beyond the scope of the current study, research should be conducted to determine whether the feedback provided through the software during individual practice can impact aural skill acquisition and error detection skills” (p. 27).

Though the number of studies is limited, technology was found to be equally as effective as traditional methods as a sight-singing assessment tool. Researchers have suggested that teachers should introduce new technologies cautiously to assist choristers in adjusting to unfamiliar assessment procedures.

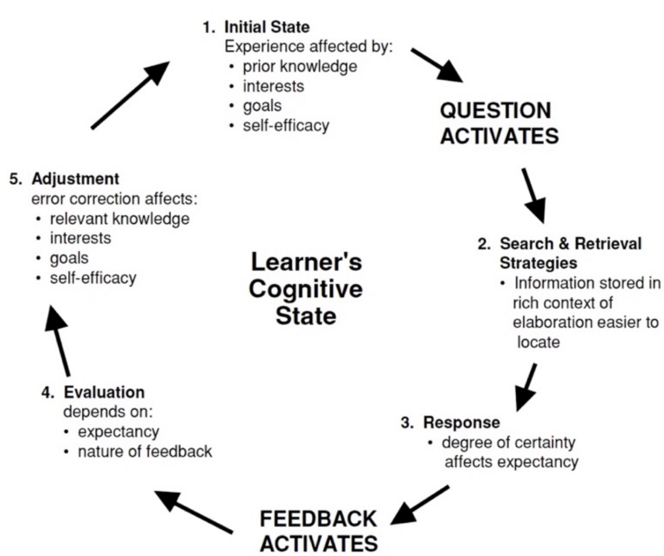
Feedback

According to Kulhavy and Wager (1993), feedback “designates any information that follows a response and allows a student to evaluate the adequacy of the response itself” (p. 3). The study of feedback has its roots in the work of E. L. Thorndike (1927; 1933), who studied how simple “right” and “wrong” feedback to student responses could affect those responses in subsequent trials. The psychologist B. F. Skinner and his study of behaviorism built on the work of Thorndike. Skinner (1965) believed that environmental stimuli either reinforced behavior or acted as punishment to diminish that behavior. Behaviorism was followed by the study of cognitivism (Gagné et al., 1981; Kulhavy & Wagner, 1993) and constructivism (Jonassen, 1990; Karagiorgi, & Symeou, 2005). Each of these learning theories offers different perspectives but have all addressed the influence of feedback on learning.

The effectiveness of feedback is dependent upon many factors (Mory, 2004). Student motivation has been found to be a factor in student attention to feedback (Timmers et al., 2013). Two factors were studied, the perceptions of the value of a task and the belief that time spent on a task would yield positive outcomes. Participants who valued a task more highly were more likely to seek feedback but did not spend more time interacting with that feedback. The learning model presented in Figure 2.1 displays complexities found when learners interacted with feedback. Two components, expectancy (students are motivated to improve), and the nature of the feedback were impactful when students were adjusting their behaviors (Bangert-Drowns et al., 1991). Learners' engagement with feedback was also affected by their acquisition and avoidance goals (Hoska, 1993). Students who viewed feedback as a means to improve, interacted with it to change their behavior, while students who viewed feedback as a measure of their ability were motivated by a fear of failure and did not engage with the feedback (Dweck, 1986).

Figure 2.1

The learner's cognitive state



Note: Adapted from Bangert-Drowns et al., 1991 by Dempsey, 1993

Despite broad interest in studying feedback in the fields of psychology and education, researchers have found that the effects of feedback were often ineffective, and occasionally had negative results (Mory, 2004). Kuhlhavy (1977) identified two conditions under which feedback was likely to produce a negative result: (a) when the information contained in the feedback was likely already known to the participant, and (b) when the task was too difficult, rendering the feedback overwhelming. Kluger and DeNisi (1996) found negative effects from feedback when it was unclear or produced an ambiguous response leading to a negative feedback loop where the participant doubted their abilities. Butler and Winne (1995) found participants became over-reliant on the provided feedback rather than learning to work independently.

A meta-analysis was conducted of 1,609 studies on the effects of feedback within a technology-based learning environment (Van der Kleij, et al., 2015). Of primary interest in this analysis was a comparison of different feedback types including, (a) *knowledge of results* (KR) (correct or incorrect response indicated), (b) *knowledge of correct response* (KCR) (correct response indicated), and (c) *elaborated feedback* (EF) (explanation provided). The effect sizes of KR and KCR feedback varied based upon the complexity of the learning task, with the higher-level outcomes having less effect. The effects of EF were found to be much more substantial but the variety of forms of EF varied among studies. Feedback, when used as reinforcement of a correct response, has not been found to affect achievement (Kuhlhavy & Wagner, 1993).

Education researchers also studied the timing of feedback on learning outcomes and retention. In many cases, the timing of feedback studied was either immediate or delayed by as much as 24 hours (Clariana, 2000; Nakata, 2015). In many cases, delayed feedback contributed significantly to memory retention over immediate feedback though students reported preferring immediate feedback over delayed. When offered the option of viewing delayed feedback, only 47% of participants chose to do so (Mullet et al., 2014).

Within music education research, the effects of learner KR has been studied (a) within teaching sequential patterns (Price, 1992), (b) during piano performance (Coffman, 1990; Banton, 1995), (c) on elementary voice development (Rutkowski & Miller, 2003; Welch, 1985), (d) while learning foreign language diction, (Steinhauer, & Grayhack, 2000), (e) on success and failure attribution (Schmidt, 1995; Vispoel, & Austin, 1993), and (f) on interval identification (Jeffries, 1967).

Real-time computerized visual feedback has been used to study singing accuracy with varying results. Graham Welch (1985) used real-time visual feedback and KR to assist elementary children in learning an echo singing task. Groups that reviewed feedback showed greater improvement than control groups. Wilson et al. (2008) and Leong and Cheng (2014) found participants who were presented concurrent visual feedback significantly improved their singing accuracy following a training period than participants who did not receive feedback. Paney and Tharp (2019) found no differences among groups in a similar study. Howard (2005) found visual feedback useful during a private voice lesson setting but cautioned against displays becoming over-complicated or ambiguous. Wilson et al. (2008) suggested using a hybrid mode of instruction where traditional methods were supplemented with technology-based visual feedback.

Despite feedback being considered an essential tool in education, results on the efficacy of feedback have been mixed. At the time of this study, no literature was found examining either the role of feedback as either KR or KCR in sight-singing achievement or the accuracy of the sight-singing feedback provided by computerized technology. The purpose of this study was to investigate the effects of feedback on sight-singing achievement, both within a sight-singing assessment session and following a series of five sessions. It also sought to compare the accuracy of the feedback available through the *SmartMusic* assessment feature when compared to an expert human rater. The following questions guided this inquiry:

1. Does the presence or timing of feedback provided by the *SmartMusic* interface affect student sight-singing achievement following a five-week treatment period?
2. Does the presence or timing of feedback provided by the *SmartMusic* interface affect student abilities to correct errors following a sight-singing attempt?
3. What is the reliability of the feedback provided by the *SmartMusic* interface when compared to human expert ratings?

Chapter 3

Method

Research Design

To isolate the effects of computerized visual feedback on sight-singing achievement, this study utilized a matched-group, repeated measure design to analyze within-session improvement and a matched pretest, posttest design to compare differences in sight-singing abilities following the treatment period. I manipulated the presence and order of the feedback provided by *SmartMusic*. Within each session of the treatment period (weeks 3-7), all participants attempted identical melodies twice. The within-session feedback group received feedback indicating correct and incorrect responses following the first attempt while the post-session feedback group received visual feedback following the second attempt. The control group received no visual feedback from the *SmartMusic* interface. Melody singing attempts were recorded twelve times from each participant over a period of nine weeks. See Figure 3.1 for a model of the research design.

Figure 3.1

Design. Matched group repeated measures with control

Group	*Week 1*	*Week 3*	*Week 4*	*Week 5*	*Week 6*	*Week 7*	*Week 9*
Within-session feedback	O_1	O_2 X O_3	O_4 X O_5	O_6 X O_7	O_8 X O_9	O_{10} X O_{11}	O_{12}
Post-session Feedback	O_1	O_2 O_3 X	O_4 O_5 X	O_6 O_7 X	O_8 O_9 X	O_{10} O_{11}X	O_{12}
Control	$O_{1.}$	O_2 O_3	O_4 O_5	O_6 O_7	O_8 O_9	O_{10} O_{11}	O_{12}

O- Melody attempt
X – Feedback provided by the *SmartMusic* interface.

Instruments

The intervention in this study was the visual feedback provided by *SmartMusic Classic* computer application loaded on an iPad Pro (10.5-inch), iOS version 12.2 (16E227) with the sight-singing instructional text, *90 Days to Sight Reading Success: A Singer's Resource for Competitive Sightsinging* by McGill and Stevens (2003). Following a sight-singing attempt, participants in the within-session feedback (WSF) and post-session feedback (PSF) groups received visual feedback from the *SmartMusic* assessment feature (see Figure 3.2) that used a proprietary voice pitch-tracking algorithm. Participants in the no-feedback/control group (NFC) sang identical excerpts with the iPad microphone turned off, negating the *SmartMusic* feedback feature. Except for the presence of feedback, the *SmartMusic* interface looked identical for all participants.

Figure 3.2.

SmartMusic interface indicating correct pitches in green and incorrect in red

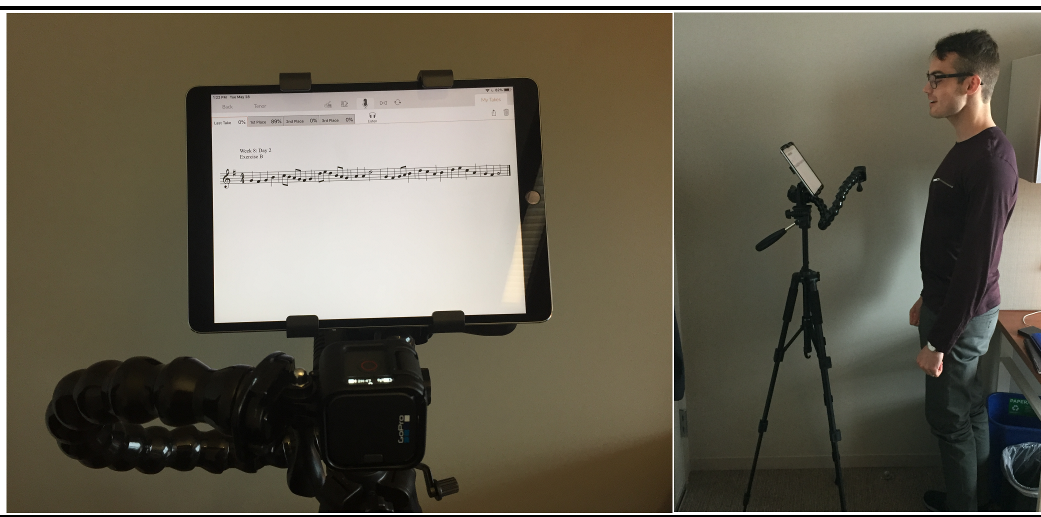
I used a GoPro HERO Session equipped with a 64GB ScanDisk Micro SD card as an audio and screen capture device for data analysis. An additional audio capture device, Zoom H4n, Handy Recorder was attempted for redundancy but was discontinued due to technical issues. At no time was any video recording made that included a participant's likeness, and audio recordings did not include participant names. I transferred all recordings to a 1TB encrypted external hard drive at the end of each day and erased data from the micro SD card.

Sight-singing samples (see Appendix 2) were selected from McGill and Stevens (2003) due to the number of possible melodies, both in quantity and variability of difficulty, and availability through the *SmartMusic* platform. All melodies were eight bars in length in 4/4 time and included the notes of the following durations: eighth, quarter, dotted quarter, and half. All exercises began and ended on tonic and were in the following keys: G major, E-flat major, F major, and D major.

The iPad was fixed on a tripod in the horizontal profile using a tablet mount within the assessment room. The tripod was placed near a wall away from any pitch source. The GoPro was attached to the tripod using a gooseneck mount far enough below the iPad to avoid obstructing participants' view, but high enough to capture the entirety of the 10.5-inch screen. Participants were allowed to adjust the height of the iPad to match their comfort. Participants who requested to sit were allowed to do so, though none asked. See figure 3.3 for two views of the research apparatus.

Figure 3.3

iPad mount with GoPro



Pilot Study

Sample procedures were piloted twice prior to running this study. In the Spring of 2018, I ran similar research procedures with Northwestern University music students (n = 12). Results of that pilot suggested the need for sight-singing melodies to be of appropriate difficulty to allow performance errors in the initial performance that could not be completely corrected in a follow-up attempt, avoiding a ceiling effect. I also piloted the sight-singing melodies used for this study with high school students who attended a summer honor choir. Using a similar procedure, I ensured that the melodies were sufficiently challenging, and the number of errors among participants was diverse. I reduced the preparation time between melodies from two minutes to 90 seconds as students' improvements diminished when the practice period was too long.

Approval by the Northwestern University Institutional Review Board (IRB) was requested and established. See Appendix B for IRB documentation. I also obtained permission for this study from participating schools at both the district and building levels. See Appendixes C and D for approval documentation. I made modifications to parental consent and student assent forms as needed until all parties granted approval.

Participants

Sampling

Participants (n = 77) for this study were a convenience sample of choral students I recruited from two high schools in my professional network. When I approached the teachers at these schools, they were eager to invite their students to engage in the research study I proposed. Both choral programs had no prior experience using technology as a sight-singing assessment tool and a teacher willing to allow students to engage in weekly sight-singing sessions during rehearsal time. Both rehearsal spaces had a practice room adjacent to the choir room that I used for individual testing during rehearsal. I purchased a one-year educators' subscription to *SmartMusic* for both teachers to allow for continued use.

Both sites were suburban high schools (grades 9-12) from a midwestern public school district. Enrollment was around 1,550 students for School A and 1,650 students for School B. Both schools had identical choral music course offerings and both used a modified block schedule. Each school had non-auditioned choirs including a tenor/bass ensemble and a treble choir. The remaining choirs were all selected by individual audition with the choir teacher. Auditioned ensembles included a select treble choir, a large mixed choir, and a small select choir, listed in order of increasing selectivity.

Recruitment

A week prior to the beginning of the nine-week study, each participating teacher granted me class time to recruit participants in person. During that recruitment, I informed students that I was a researcher from Northwestern University studying technology and the individual assessment of sight-singing. I informed prospective participants that individual assessment had been found to improve sight-singing achievement (Demorest, 1998) and that I hoped they would improve sight-singing skills through participation in this study. Parent consent and student assent forms were distributed and signed forms were collected by the participating teachers. At my request, the participating teachers sent an email to all parents informing them about the study and encouraging participation. I collected all completed forms from the teacher a week later and each participant was assigned a research code.

Final Sample

Participants for this study represented a convenience sample. I received 42 completed consent forms from School A resulting in a 25% participation rate at that site. Forty completed forms were returned from School B for a 26% participation rate. Among those who submitted completed consent forms, all completed the pretest procedures. Two students withdrew from the study immediately after the pretest and two more were unable to complete the study due to low attendance, establishing an attrition rate below 5%. Participants who withdrew or failed to complete the entire sequence had their data deleted. Thirty-nine participants completed the study from School A and 38 from school B. See Table 3.1 for a breakdown of participants by select background variables.

Table 3.1

Student Participation by School, Disaggregated by Vocal Range, and Choral Ensemble

Choir	School A	School B
Tenor/Bass	8	6
Treble Choir	10	6
Select Treble	4	6
Mixed Large Ensemble	10	9
Mixed Select Ensemble	7	11
Vocal Range		
Tenor/Bass	13	13
Soprano/Alto	26	25
Total (n = 77)	39	38

Procedures

Pretest

Henry (2014) suggested the importance of building familiarity with the *SmartMusic* interface to improve student comfort with the assessment process. Therefore, the week of the pretest, I introduced the *SmartMusic* interface through group sight-singing practice. I mirrored my iPad on the classroom screen using a projector during the warm-up time at the beginning of a rehearsal. I also presented a list of strategies used by effective sight-singers during practice (Killian & Henry, 2005) that included (a) tonicize the key, (b) practice out loud, (c) work on problem areas, (d) physically keep the beat, and (e) keep the beat steady. All participants sang a sample melody using the interface prior to taking the pretest. Students who were absent for the group presentation were given an identical presentation in person before taking the pretest.

The pretest began a nine-week research period that allowed for (a) a week for pretest assessments, (b) a week off, (c) five weeks of once-weekly assessment sessions, (d) another week off, and (e) a week for posttest assessments. Using an alternating pretest/posttest design, participants sang one of two randomly assigned melodies for the pretest. An equal number of participants sang Melody A for the pretest as sang Melody B. See Appendix F for the pretest melodies.

Stimulus

The pretest was administered by playing a screen-capture video of the *SmartMusic* interface that included a tonic triad (*d m s m d s, d*), thirty seconds of participant self-guided practice, another tonic triad (*d m s m d s, d*), a four-beat count off, and 50 seconds to complete the melody. The entire pretest stimulus ran for approximately one minute and 50 seconds. A click-track, quarter note indicator, and visual feedback, common features of the *SmartMusic* interface, were not enabled during the pretest. See figure 3.4 for screenshots from the pretest interface.

Figure 3.4

Pretest/Posttest Interface

Initial interface *During practice* *While singing*

Pretest Administration

To administer the pretest, I wrote the last names of participating students in random order on a whiteboard at the front of each room. Students were instructed to report to the sight-singing room when the name above theirs was erased. Before they entered the assessment room, I reminded the participants of the features of the stimulus, including tonic triad and practice time, and informed them they would have 50 seconds to complete the melody. Participants were asked to sing the melody once without going back or starting over to achieve the best possible score.

Participants were then invited into the assessment room (a standard practice room at both schools) and were asked to press the play icon on the iPad when I left the room and closed the door. Using a wireless connection from the GoPro to an out-of-service iPhone, I was able to record and visually monitor the GoPro video. At the end of the pretest video stimulus, I opened the door, and thanked the participant for their time. After they left, I reset the apparatus for the next participant.

Following the completion of the pretest, I scored each sight-singing attempt for rhythm and pitch accuracy and noted the students' vocal range (tenor/bass or treble). I created matched groups by ranking composite (pitch plus rhythms) scores by melody and school, establishing four ranked lists. Then I assigned participants, in order, to one of three conditions: (a) within-session feedback (WSF), (b) post-session feedback, and (c) no-feedback/control (NFC). See Table 3.2 for a list of matched-group pretest scores.

Table 3.2

Comparison of Matched Group Pretest Scores

Melody and Group	*n*	*M*(*SD*)	95% CI
Melody A			
WSF group	12	.234(.081)	[.057, .412]
PSF	13	.264(.084)	[.081, .447]
NFC	14	.214(.072)	[.057, .371]
Melody B			
WSF	14	.388(.090)	[.195, .581]
PSF	11	.332(.103)	[.103, .561]
NFC	14	.308(.081)	[.113, .483]

Note. All groups had 13 or 14 participants prior to attrition ($n = 4$).

Weekly sight-singing session Administration

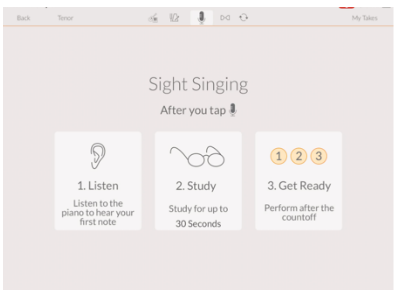
A series of five, once-weekly assessment sessions began on the third week of the study. As with the pretest, I presented a sight-singing melody to each class using the *SmartMusic* interface and a projector. The purpose of doing this was to familiarize participants with the click track and quarter-note indicator, features not enabled during the pretest. Students who were absent during the first day of Week 1 were able to practice the same prior to their first assessment session. I did not discuss the existence of research groups or the feedback features of the *SmartMusic* interface with participants. The physical setup of the assessment room was identical to that used during the pretest. As with the pretest, I also wrote the last names of participants on a whiteboard in the order they were to perform. The order of participants was the same as pretest and this order remained the same throughout the remainder of the study.

Prior to participants entering the room, I reset the *SmartMusic* interface and deleted any existing attempts from the iPad. I placed a sticky note which included each participant's research code in the corner of the iPad masking the percentage score from the *SmartMusic* interface. I also set the apparatus to either the tenor/bass or treble range based upon the participants' selected octave during the pretest. I ensured the microphone feature was enabled for participants in the within-session feedback group and turned off for participants in both the post-session feedback and control groups, allowing feedback when enabled and eliminating it when disabled. As students entered the assessment room, I verbally reviewed the assessment procedures that were as follows: (a) when I exit the room, press the microphone icon on the *SmartMusic* interface, (b) this will begin a 30 second practice period that will be preceded and followed by the tonic triad (*d m s m d s, d*), (c) sing the melody while keeping up with the click-track and quarter note indicator, (d) after completing the melody, take 60 seconds to review the melody and try to correct any errors, (e) I will re-enter the room and reset the apparatus for a second attempt, (f) when I exit the room press the microphone icon on the *SmartMusic* interface, (g) sing the melody a second time and try to improve upon your initial attempt, and (h) exit the room. When I entered the room following the first attempt and 60 seconds of practice, I enabled the microphone feature for participants in the post-session feedback group and disabled it for those in the within-session feedback and left it disabled for those in the NFC group. Each session took approximately four minutes and 35 seconds. Figure 3.5 displays screenshots of the weekly session interface.

Figure 3.5

Weekly session interface

Initial interface



During practice

While singing

Posttest procedures

I administered the posttest during the ninth week of the study following a one-week hiatus after the fifth weekly assessment session. Posttest procedures were identical to those used during the pretest, except participants sang the melody they did not sing during the pretest. As with the pretest, I presented the *SmartMusic* interface on a screen, and each class practiced sight-singing a melody without the use of a click-track or quarter note indicator.

Post-research Debrief

My planned return to the participating schools to discuss the findings with participants was made impossible due to the COVID-19 pandemic. My visit to debrief participants was rescheduled for a time when in-person instruction was possible.

Data Analysis

I scored all pretest and posttest melody attempts (n = 154) using method I piloted previously. A random sample of pilot data revealed a 90.7% agreement of scores when comparing agreements divided by agreements plus disagreements (C. K. Madsen & C. H. Madsen, 1970) between two raters. I employed the following procedures: Each eight-measure sight-singing sample was divided into two, two-count chunks (counts 1-2 and counts 3-4) for a total of 16 chunks. Each chunk was then awarded one point for the correct notes and one point for correct rhythm for a total of 32 possible points per sample. If any error was made within a chunk, the entire chunk was awarded a zero. See appendix A for a complete list of pretest/posttest scoring guidelines. All scores for both pitch and melody were converted into a proportion of correct chunks per attempt. A random sample of approximately 20% (n = 30) pretest and posttest melodies were scored by an additional expert rater to establish reliability. A proportion of agreements divided by agreements plus disagreements (C. K. Madsen & C. H. Madsen, 1970), yielded a proportion of agreement of (.925) for pitch and (.856) for rhythm.

I also scored all weekly assessment session attempts (n = 770). In order to account for the unique nature of the *SmartMusic* interface, I used a different scoring method than I used for the pretest. I awarded a single point to each correct pitch and a point for each correct rhythm similar to other studies (Henry 2004; 2011). Unique to this study, however, participants were required to stay within a quarter step, sharp or flat, of the written pitch and rhythms were required to be aligned with the click track and quarter note indicator to be marked correct. See Appendix A for a complete list of scoring guidelines. In order to ascertain reliability, an additional expert rater scored 30% (n = 235) of the sight-singing attempts, selected at random. Using a formula of agreements divided by agreements plus disagreements, I was able to determine a proportion of agreement for pitch (.908) and rhythm (.852) for the melodies in weeks one through five.

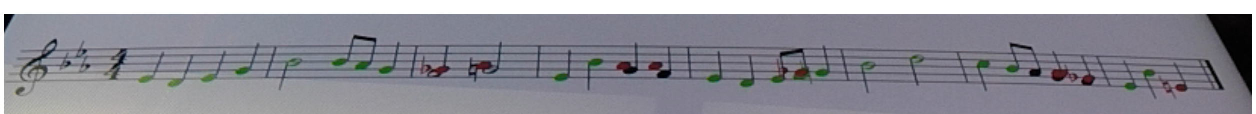
In order to determine the reliability of my pitch and rhythm proportion scoring, I ran a Type A (absolute agreement), intraclass correlation coefficient (ICC), comparing my scores to the additional rater's scores revealing a high degree of reliability. The single measures ICC for pitch scores was .939 with a 95% confidence interval from .923 to .953, $F(244, 244) = 32.337$, $p < .001$. The single measures ICC for rhythm was .904 with a 95% confidence interval from .569 to .959, $F(244, 244) = 33.349$, $p < .001$. As the ICC for both pitch and rhythm fell within the "*excellent reliability*" range (Koo & Li, 2016), I proceeded to use my full set of scores without modification.

SmartMusic Assessment Reliability

In order to address the reliability of the *SmartMusic* interface compared to human raters, I compared the visual feedback captured by the video apparatus to my scores. Because the apparatus did not account for the possibility of separate pitch and rhythm scores, I considered a note correct when I scored both the pitch and rhythm as correct. I considered the note correct from the *SmartMusic* interface when the notehead was green instead of black or red. I ignored all other extraneous marks on the feedback. See Figure 3.6 for a comparison of human and computerized scoring.

Figure 3.6

The SmartMusic Assessment Feature Interface



Human scoring	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
Agreements	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0

Note: Agreements (23) were divided by agreements plus disagreements (23+6) for .793 for this attempt

Summary

The purpose of this study was to examine the effects of computerized feedback on secondary choristers' sight-singing and melody singing achievement. In order to address the first research question, I compared pretest and posttest sight-singing scores from participants in three groups using two different melodies. To address the second research question, I compared the scores from participants' initial sight-singing attempt to a follow-up attempt on the same melody following a period of practice. I addressed the third research question by comparing human scores of a melody singing attempt to the visual feedback provided by the *SmartMusic* interface of the same attempt. The following chapter will address the findings of this study.

Chapter 4

Results

The purpose of this study was to investigate the effects of the presence and timing of computerized feedback on sight-singing achievement, both within a five-minute sight-singing assessment session and following a five-week treatment period. I also sought to study the reliability of the visual feedback available through *SmartMusic* using an iPad interface when compared to a human rater. In this chapter, I will detail the statistical procedures I used and the findings from each research question.

Research Question 1

For the first research question, I sought to find if either the order or presentation of computerized feedback affected sight-singing achievement following a five-week treatment period. I randomly assigned participants to sing one of two pretest/posttest melodies, A or B, and they were placed in the following matched groups: (a) within-session feedback (WSF), (b) post-session feedback (PSF), and (c) no feedback/control (NFC). This section will detail the statistical procedures I used to compare research groups following a five-week treatment period and present the results. I will discuss the findings in terms of composite scores, pitch-only scores, and rhythm-only scores.

Composite Scores

To compare possible growth in sight-singing scores among groups, I performed a two-by-two repeated measures ANOVA of pretest and posttest pitch and rhythm scores with condition as a between-subjects factor. As the assumption of sphericity could not be met, a Greenhouse-Geisser correction was applied, revealing no significant differences between pretest and posttest composite scores F(1.000, 71.000) = 2.106, p = .151, η_p^2 = .029. Additionally, between-subjects comparisons revealed no significant differences between groups F(2, 71,) = 2.492, p = .090, η_p^2 = .066. See Table 4.1 for a comparison of pretest and posttest means by group and melody.

Table 4.1

Pretest/posttest composite score comparison by melody and condition

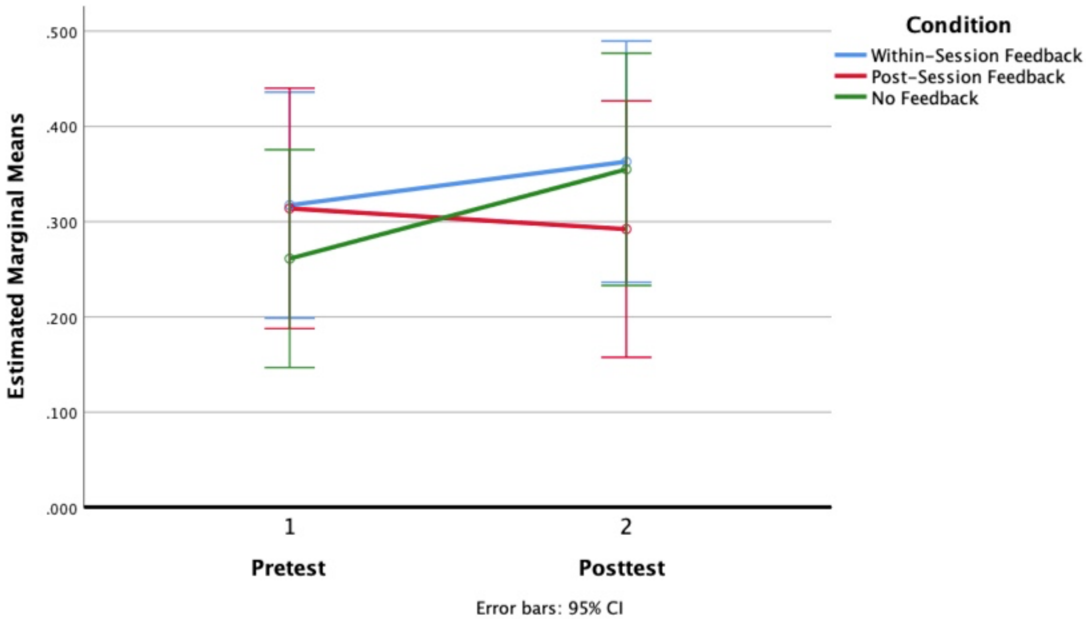
Composite score	Pretest		Posttest	
	M	*SD*	*M*	*SD*
Within-session Feedback				
Melody A	.264	.085	.385	.089
Melody B	.385	.089	.341	.089
Post-session Feedback				
Melody A	.277	.085	.216	.102
Melody B	.362	.097	.351	.089
No feedback/control				
Melody A	.214	.082	.259	.086
Melody B	.308	.082	.451	.086

Note: Composite scores are reported as proportion correct notes and rhythms

A comparison of composite gains from pre- to posttest by condition revealed positive gains in the within-session feedback group (ΔM = .046) and no-feedback group (ΔM = .094) but negative gains in the post-session feedback group (ΔM = -.036). None of these differences were statistically significant. See Figure 4.1 for a comparison of pretest and posttest scores by group.

Figure 4.1

Comparison of pretest and posttest scores by group



Note: Composite scores are reported as proportion of accurate pitches and rhythms

Pitch Scores

A repeated measures ANOVA was conducted to ascertain if condition made differences in pitch scores between the pretest and posttest. Using a Greenhouse-Geisser correction, no significant differences were found between pretest and posttest pitch scores $F(1.000, 71.000) = 2.396$, $p = .126$, $\eta_p^2 = .033$. See Table 4.2 for a complete comparison of pitch means by condition and melody.

Table 4.2

Pretest/posttest pitch score comparison by melody and condition

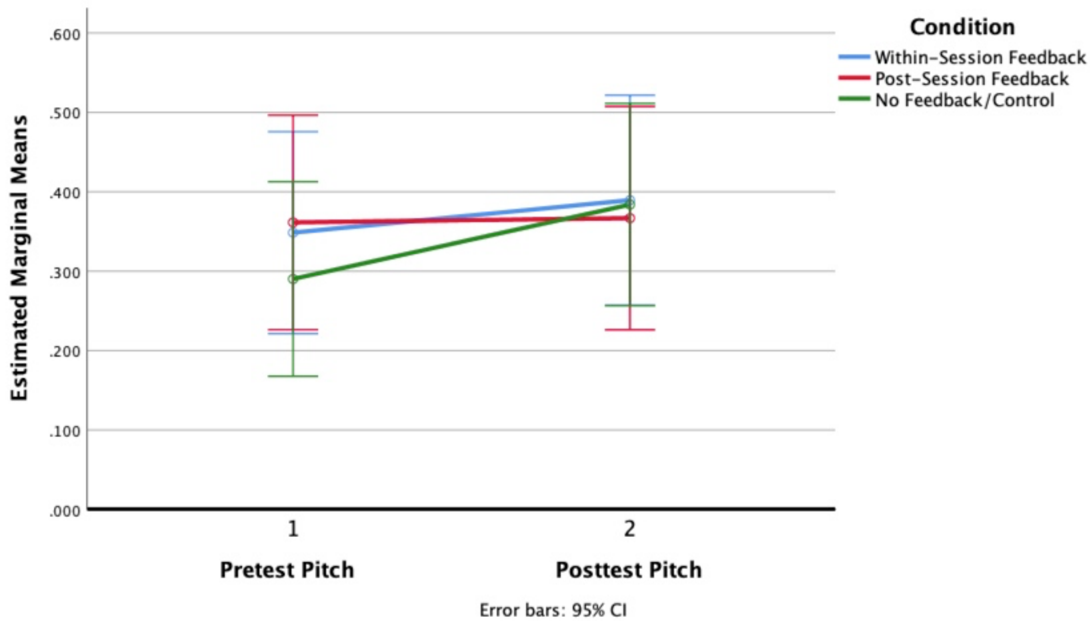
Pitch only score	Pretest *M*	Pretest *SD*	Posttest *M*	Posttest *SD*
Within-session Feedback				
Melody A	.255	.090	.413	.092
Melody B	.365	.090	.365	.092
Post-session Feedback				
Melody A	.317	.090	.281	.105
Melody B	.281	.105	.433	.092
No feedback				
Melody A	.232	.087	.254	.089
Melody B	.348	.087	.513	.089

Note: Scores are reported as a proportion of correct pitches.

A comparison of pitch gains from pre- to posttest by condition reveals similar results to that of the composite comparison. Positive gains were found in the within-session feedback group ($\Delta M = .050$) and no-feedback group ($\Delta M = .094$) but revealed negative gains in the post-session feedback group ($\Delta M = -.011$). None of these differences were statistically significant. See figure 4.2 for a comparison of pretest and posttest pitch scores by condition.

Figure 4.2

Comparison of pretest and posttest pitch scores by condition



Note: Composite scores are reported as proportion of accurate pitches and rhythms

A Greenhouse-Geisser corrected repeated measures ANOVA found no significant differences in rhythm scores between pretest and posttest attempts F(1.000, 71.000) = .903, p = .345, η_p^2 = .013. See Table 4.3 for a list of rhythm scores by melody and condition.

Table 4.3

Pretest/posttest rhythm score comparison by melody and condition

Rhythm only score	Pretest		Posttest	
	M	*SD*	*M*	*SD*
Within-session Feedback				
Melody A	.274	.091	.356	.096
Melody B	.298	.091	.317	.096
Post-session Feedback				
Melody A	.250	.091	.150	.109
Melody B	.306	.104	.269	.096
No feedback				
Melody A	.196	.088	.254	.093
Melody B	.268	.088	.406	.093

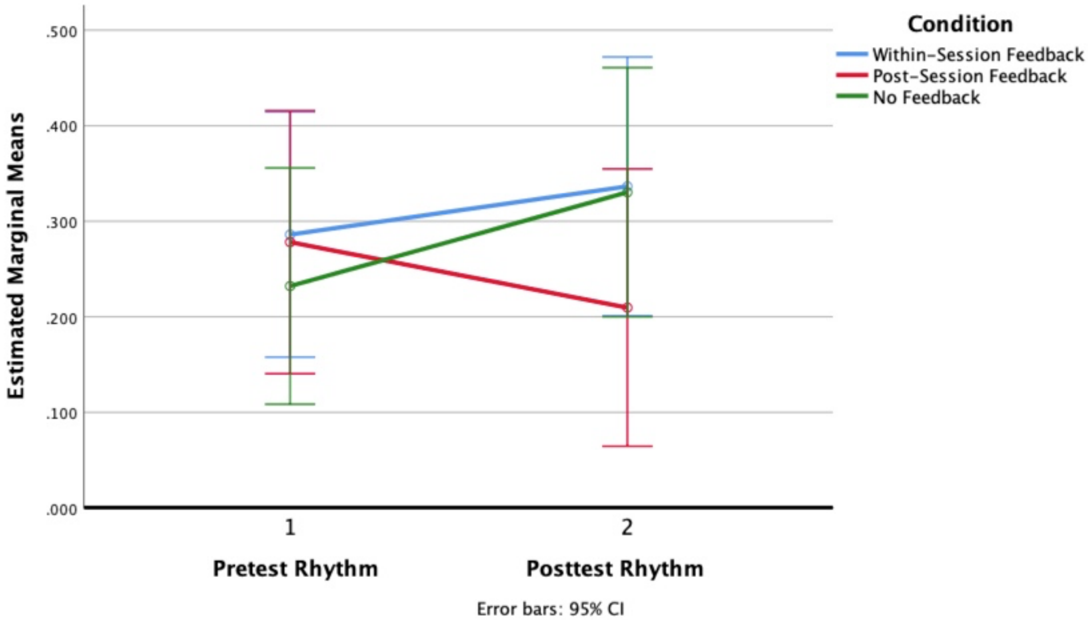
Note: Scores are reported in proportion of correct rhythms

Rhythm Scores

A comparison of rhythm gains from pretest to posttest reveals a greater drop in scores among those in the post-session feedback group ($\Delta M = -.068$) than those indicated by the same group in composite and pitch scores. Participants in the within-session feedback group ($\Delta M = .051$) and no feedback group ($\Delta M = .098$) showed positive gains. None of these differences were statistically significant. See Figure 4.3 for a comparison of pretest and posttest scores by rhythm.

Figure 4.3

Comparison of pretest and posttest rhythm scores by condition



Note: Composite scores are reported as proportion of accurate pitches and rhythms

I compared initial attempts from each of the once-weekly sight-singing assessment sessions to track the possibility of individual improvement. Results from a five-session repeated measures ANOVA of participants' initial weekly attempts revealed a significant main effect F(4, 292) = 33.637, $p < .001$, $\eta_p^2 = .315$. Pairwise post-hoc analysis indicated significant differences by attempt with the melody in week 5, producing the highest mean score. See Table 4.4 for pairwise comparisons of changes in initial composite attempt scores.

Table 4.4

Pairwise comparisons of differences in initial composite attempt scores

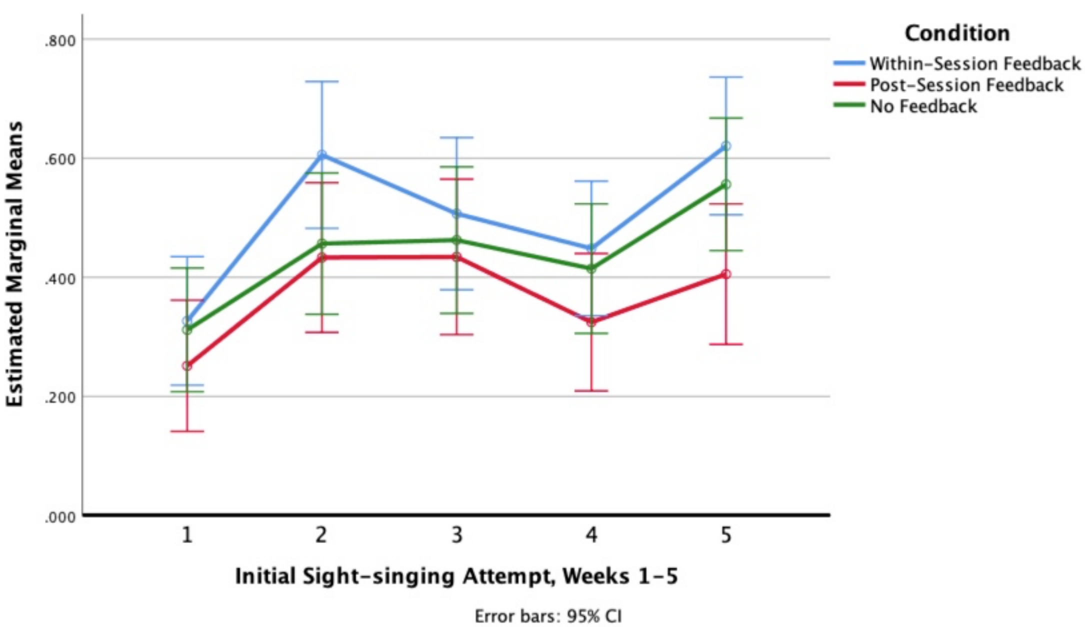
Week	2	3	4	5
1	.202**	.171**	.099**	.231**
2		-.031	-.103**	.029
3			-.072*	.059
4				.131**

Note: *$p < .05$. **$p < .001$.
Scores are reported as differences proportion of correct pitches and rhythms

Though participants in the WSF group produced the highest mean composite score during the initial attempt, differences among groups were not significant. See Figure 4.4 for a comparison of initial sight-singing attempts by group.

Figure 4.4

Comparison of Initial Sight-singing Attempts by condition



Note: Composite scores are reported as proportion of accurate pitches and rhythms

Research Question 2

With question two, I sought to ascertain if the presence or timing of feedback affected participants' ability to improve accuracy on the performance of a melody following an initial sight-singing attempt. This section will first examine the overall effects of three conditions on within-session improvement by analyzing all five weeks concurrently, followed by an analysis of improvement within each of the five weekly sessions. This section explores the factors that predicted success on the second (follow-up) attempt during each of the five sessions.

Weeks 1-5

The following three-level panel data regression model was developed, regressing several independent variables on the follow-up attempt composite score. The first level model included follow-up attempt composite scores as the outcome variable, and the initial attempt composite scores as a predictor variable:

$$FollowupAttemptComp_i = \alpha + \beta_1 InitialAttemptComp_i + \varepsilon$$

The second model added dummy variables for the within-session feedback group and the post-session feedback group:

$$FollowupAttemptComp_i = \alpha + \beta_1 InitialAttemptComp_i + \beta_2 WSF_i + \beta_3 PSF_i + \varepsilon$$

The final model added other dichotomous predictor variables, including choir selection, school, and voice range:

$$FollowupAttemptComp_i = \alpha + \beta_1 InitialAttemptComp_i + \beta_2 WSF_i + \beta_3 PSF_i + \beta_4 STC_i + \beta_5 LMC_i + \beta_6 SMC_i + \beta_7 School_i + \beta_8 Voice_i + \varepsilon$$

In all three models, the initial attempt was the strongest significant predictor of success on the follow-up attempt. Other factors were also significant, including choir selection and school attendance. Research condition and voice range were not significant predictors of the outcome. See Table 4.5 for the regression analysis.

Table 4.5

Summary of Hierarchical Regression Analysis for Variables Predicting Follow-up Melody Accuracy (n= 384)

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Constant	.121	.014		.132	.017		.132	.023	
Initial attempt	.891	.026	.871**	.895	.026	.874**	.815	.031	.796**
WSF[a]				-.030	.019	-.044	-.028	.019	-.042
PSF[b]				-.012	.019	-.018	-.010	.019	-.014
Select Treble Choir[c]							.027	.027	.084*
Large Mixed Choir[d]							.080	.021	.108**
Small Mixed Choir[e]							.088	.024	.120**
School[f]							-.041	.016	-.064*
Treble Voice[g]							.006	.017	.010
R^2		.757			0.760			.777	
F for change in R^2					1.411			5.97*	

Note:
[a]Within-session feedback condition = 1, Post-session feedback and control = 0.
[b]Post-session feedback condition = 1, Within-session feedback and control = 0.
[a, b]Control group is constant = 0
[c, d, e]Non-auditioned choir is constant = 0
[c]Treble Choir= 1, [d]Large Mixed Choir = 1, [e]Small Mixed Choir = 1
[f]School A = 1, School B = 0
[g]Treble Voice = 1, Tenor/Bass Voice = 0
$*p < .05$. $**p < .001$.

In order to determine any effect of variables on possible gains between attempts, the following two-level regression model was created. See Table 4.6 for the regression results. The first model included composite gain scores as the outcome variable with dummy variables representing within-session feedback and post-session feedback groups:

$$GainScoreComp_i = \beta_1 WSF_i + \beta_2 PSF_i + \varepsilon$$

The second model added the remaining dichotomous predictor variables, including choir selection, school, and voice range:

$$GainScoreComp_i = \beta_1 WSF_i + \beta_2 PSF_i + \beta_3 STC_i + \beta_4 LMC_i + \beta_5 SMC_i + \beta_6 School_i + \beta_7 Voice_i + \varepsilon$$

Table 4.6

Summary of Hierarchical Regression Analysis for Variables Predicting Gains in Composite Scores Between Initial and Follow-up Attempts (n = 384)

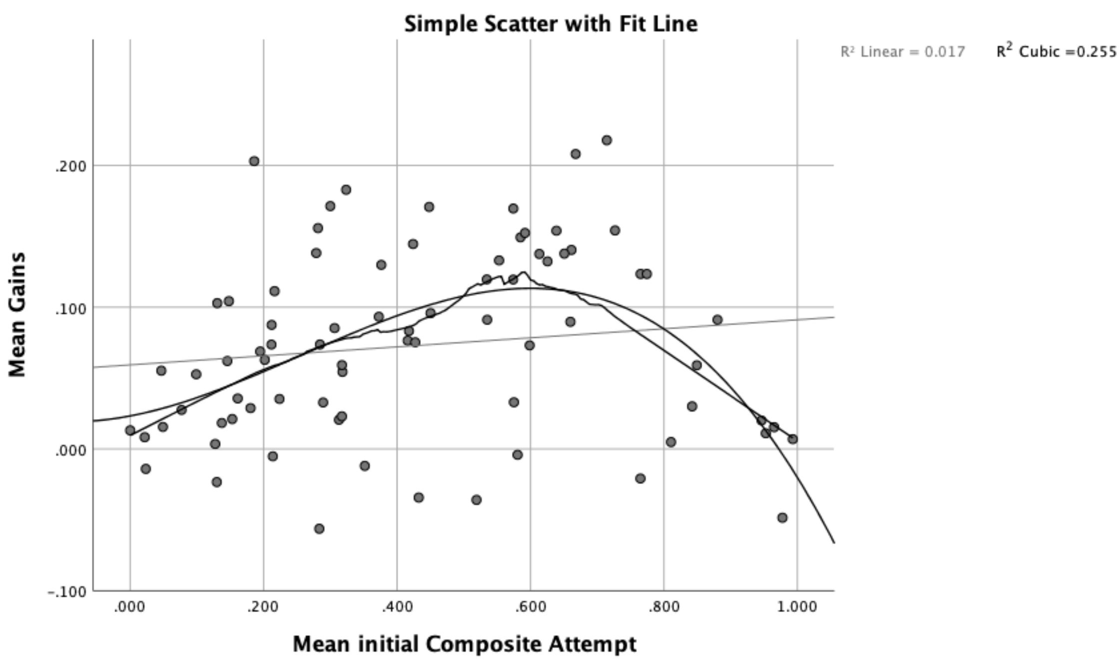
	Model 1			Model 2		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Constant	.087	.013		.083	.022	
WSF[a]	-.038	.019	-.115*	-.040	.020	-.118*
PSF[b]	-.005	.020	-.015	-.007	.020	-.019
Select Treble Choir[c]				.030	.027	.063
Large Mixed Choir[d]				.042	.021	.112*
Small Mixed Choir[e]				.010	.021	.027
School[f]				-.031	.016	-.099
Treble Voice[g]				.007	.018	.021
R^2		.012			.035	
F for change in R^2					0.133	

Note:
[a]Within-session feedback condition = 1, Post-session feedback and control = 0.
[b]Post-session feedback condition = 1, Within-session feedback and control = 0.
[c, d, e]Non-auditioned Choir is constant = 0
[c]Treble Choir= 1, [d]Large Mixed Choir = 1, [e]Small Mixed Choir = 1
[f]School A = 1, School B = 0
[g]Treble Voice = 1, Tenor/Bass Voice = 0
*$p < .05$. **$p < .001$.

I applied a Pearson's *r* correlation between mean initial composite scores and mean composite gains to determine if a relationship existed between how well participants scored on their first attempt and how much improved during their second attempt. A weak non-significant positive relationship was found $r = .130$, $p = .258$. When I applied a cubic line-of-fit to a scatterplot, Figure 4.5, comparing average initial attempts and average gains among all participants, an inverted-U shaped line was revealed ($R^2 = .255$) that better accounted for variance in the data than a linear line ($R^2 = .017$).

Figure 4.5

Scatterplot of Average Initial Attempt Score and Average Gains



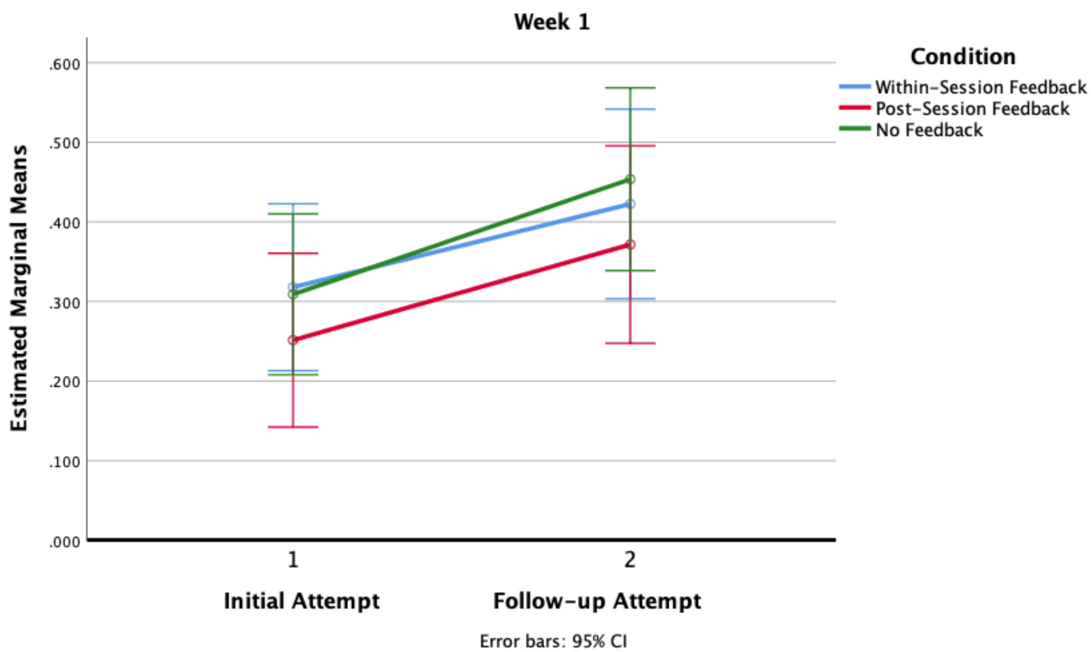
Note: Composite and gain scores are reported in proportion of correct notes and rhythms

Week 1

In order to determine if differences in gains changed by group over time, I compared scores from each week individually. I ran a two-way repeated measures ANOVA of scores comparing the pitch and rhythm scores from an initial attempt and pitch and rhythm scores from a follow-up attempt. During week one, participants scored significantly higher on the follow-up (M = .416 SD = .035) attempt than they did during the initial (M = 293, SD = .030) attempt $F(1.00, 75.000) = 51.618$, $p < .001$, $\eta_p^2 = .408$ using Greenhouse-Geisser corrected degrees of freedom. Differences among groups were non-significant and showed nearly parallel improvement across groups.

Figure 4.6

Comparison of Week 1Initial Attempt Scores and Follow-up Scores by Condition

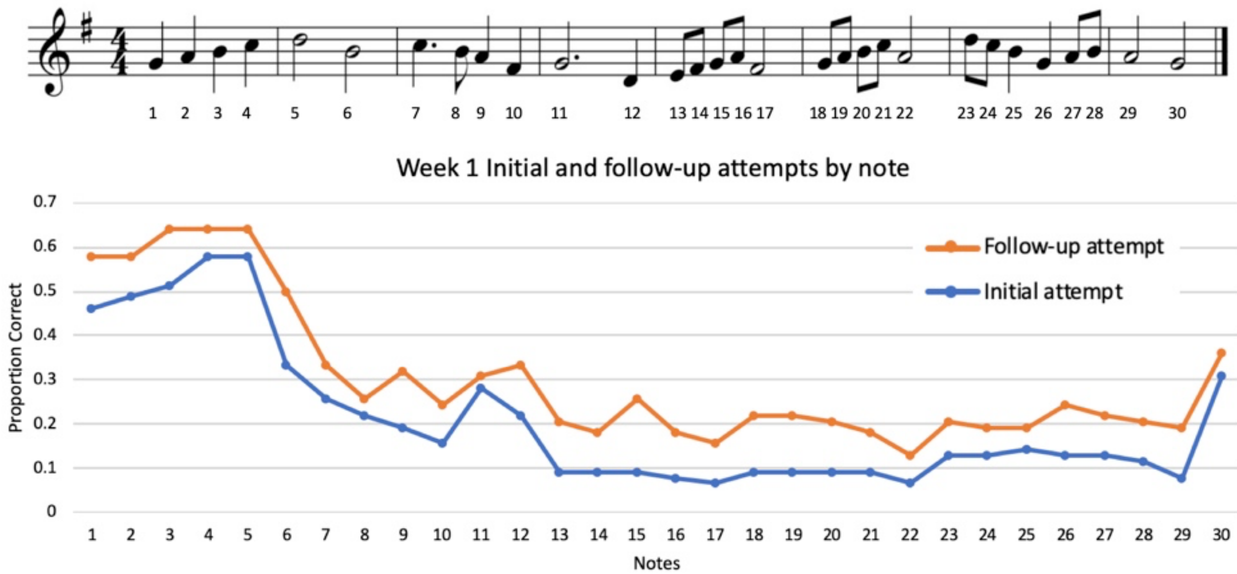


Note: Composite scores are reported as proportion of accurate pitches and rhythms

During Week 1, initial pitch mean scores (M = .318, SD = .034) were higher than initial mean rhythm scores (M = .268, SD = .032). The same was true for the follow-up attempt as mean pitch scores (M = .430, SD = .039) were higher than mean rhythm scores (M = .402, SD = .037) however, rhythm scores showed more gains (ΔM = .134) than pitch scores (ΔM = .112). The greatest gains in pitch occurred in the final three notes of the seventh measure. The greatest rhythm gains were at the beginning of measure 5. See Figure 4.7 for a display of the composite proportion correct by note during the initial and follow-up attempt.

Figure 4.7

Proportion Correct by Note During the Initial and Follow-up Attempts

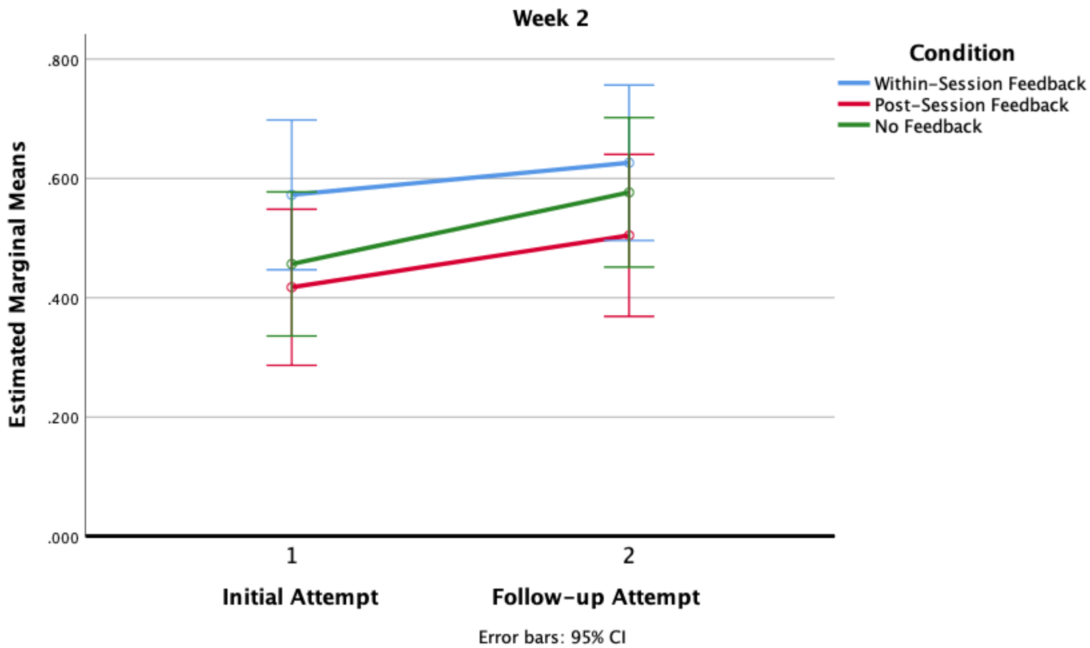


Week 2

I compared Week 2 initial pitch and rhythm scores to pitch and rhythm scores from the follow-up attempt using a two-way repeated measures ANOVA. As with Week 1, pitch and rhythm scores were significantly higher during the follow-up attempt $F(1.00, 72.000) = 26.477$, $p < .000$, $\eta_p^2 = .269$. Pairwise post-hoc analysis found no significant differences between initial and follow-up attempts by condition.

Figure 4.8

Comparison of Week 2 Initial Attempt Scores and Follow-up Scores by Condition

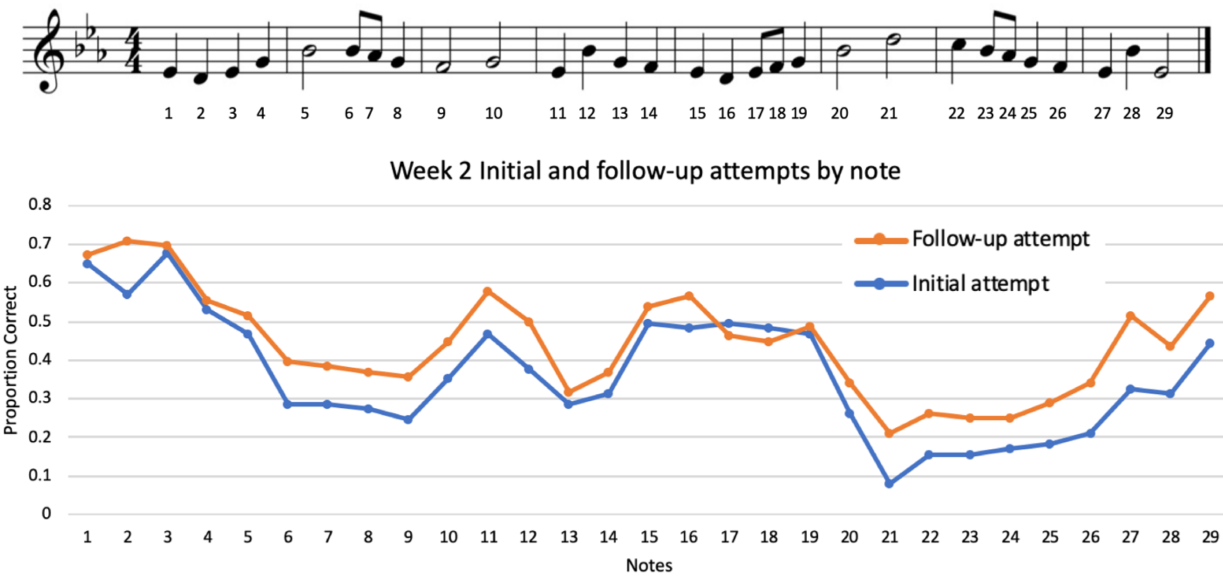


Note: Composite scores are reported as proportion of accurate pitches and rhythms

While each group showed significant gains, these gains were not as pronounced as those during Week 1. However, the initial attempt scores during Week 2 were significantly higher than the initial scores during Week 1. Additionally, differences between pitch and rhythm scores were statistically significant F(1, 72.000) = 4.963, p = .029, = η_p^2 =.065. Also in contrast to Week 1, Week 2 results found initial rhythm scores (M = .498, SD = .039) were higher than initial pitch scores (M = .467, SD = .039). During the follow-up attempt, means were higher for rhythm (M = .609, SD = .040) than pitch (M = .530, SD = .041). The resulting differences between initial and follow-up scores showed that rhythm showed greater gains than pitch scores. The greatest gains in pitch occurred during the first note of the fourth measure. The greatest rhythm gain occurred in the second note of the last measure. See Figure 4.9 for a display of the proportion correct by note during the initial and follow-up attempts.

Figure 4.9

Proportion Correct by Note During the Initial and Follow-up Attempts

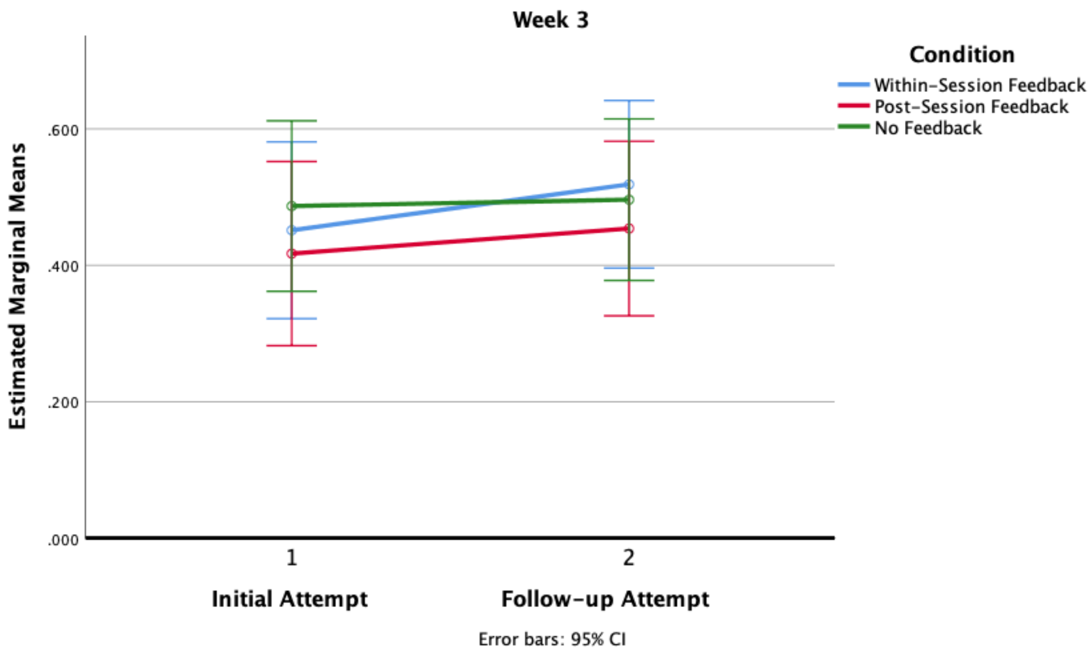


Week 3

A two-way repeated measures ANOVA was used to compare Week 3 initial pitch and rhythm scores to follow-up attempts. Unlike weeks 1, 2, 4, and 5, Results found no main effect between the initial and follow-up attempt F(1.000, 75.000) = 2.861, p = .095, η_p^2 = .037. Rhythm scores were slightly higher than pitch scores during both the initial and follow-up attempts. These differences were non-significant.

Figure 4.10

Comparison of Week 3 initial attempt scores and follow-up scores by condition

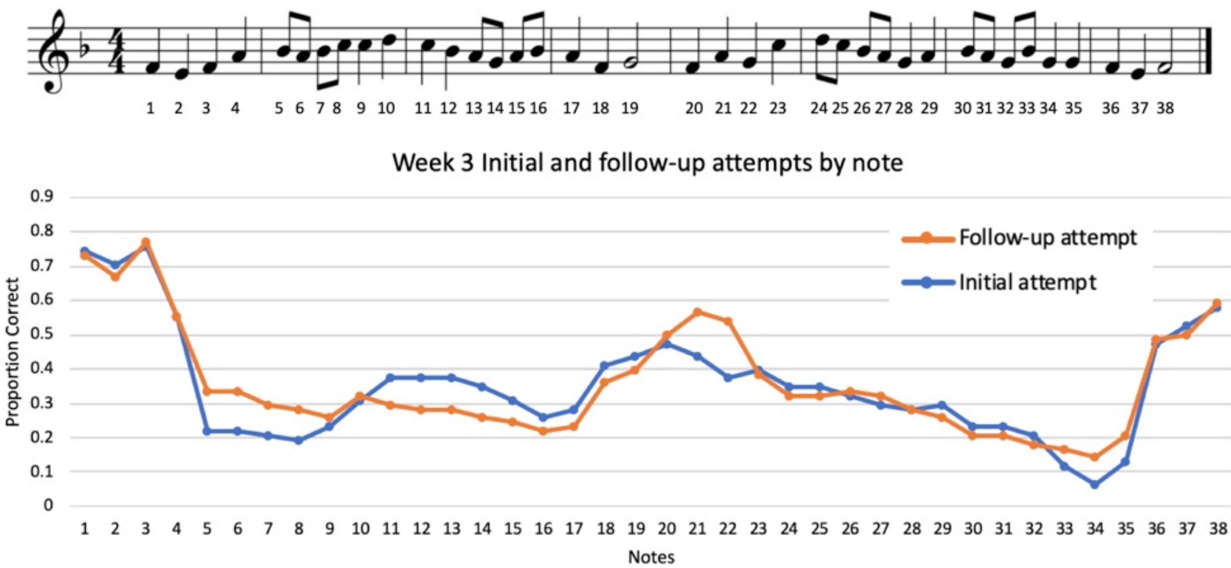


Note: Composite scores are reported as proportion of accurate pitches and rhythms

Rhythm scores showed slightly more gains ($\Delta M = .051$) than pitch scores ($\Delta M = .026$). It may be worth noting that the first measure of Week 3 exhibited higher initial pitch scores than any other measure during all five weeks of the treatment period. Gains during this measure, however, were negligible or negative. See Figure 4.11 for a display of the proportion correct by note during the initial and follow-up attempts.

Figure 4.11

Proportion Correct by Note During the Initial and Follow-up Attempts

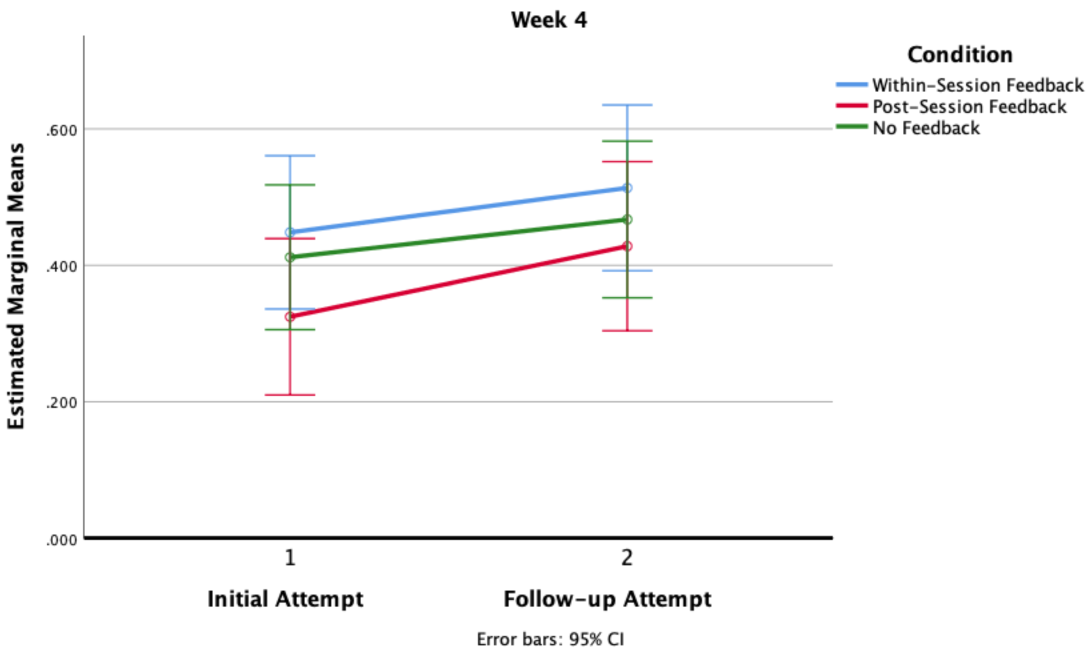


Week 4

A two-way repeated measures ANOVA was used to compare initial pitch and rhythm scores to follow-up pitch and rhythm scores. Results indicate that follow-up scores were significantly higher than those during the initial attempt $F(1.000, 74.000) = 16.665$, $p < .001$, $\eta_p^2 = .184$. A post-hoc pairwise comparison found no differences among groups, however.

Figure 4.12

Comparison of Week 4 initial attempt scores and follow-up scores by condition

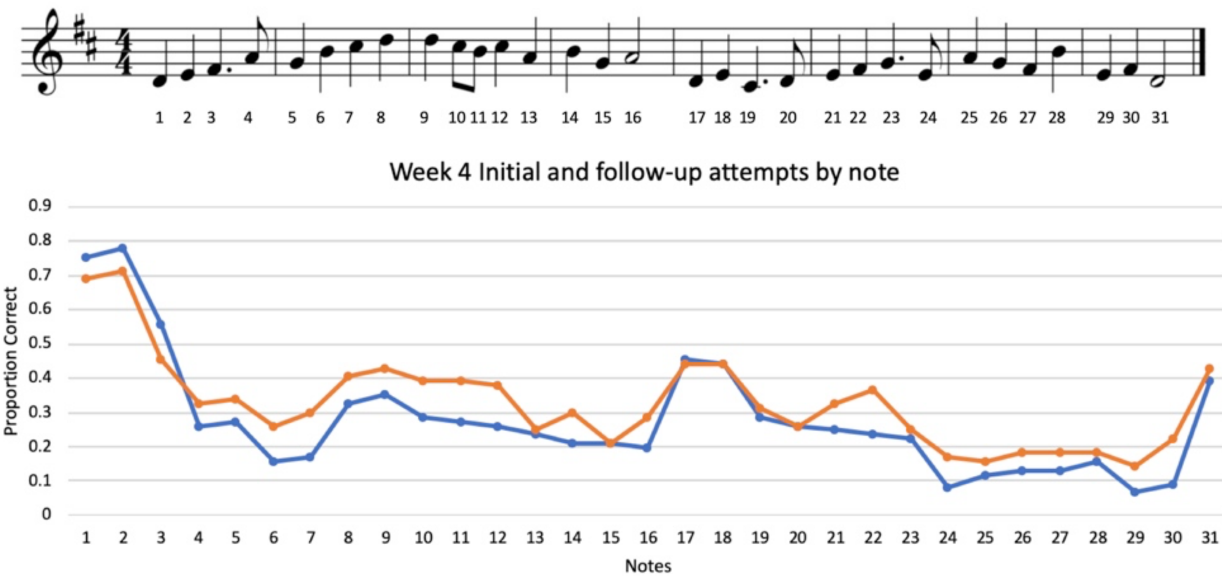


Note: Composite scores are reported as proportion of accurate pitches and rhythms

Similar to weeks 2, rhythm scores were significantly higher than pitch scores $F(1.000, 74.000) = 4.015$, $p = .049$, $\eta_p^2 = .051$. Gains between pitch ($\Delta M = .079$) and rhythm ($\Delta M = .070$) were nearly identical. The highest proportion of correct note scores (.792) during the five-week treatment occurred with the second note of the first measure. The greatest gains among notes occurred in third note of the second measure. See Figure 4.13 for a display of the proportion correct by note during the initial and follow-up attempts.

Figure 4.13

Proportion Correct by Note During the Initial and Follow-up Attempts

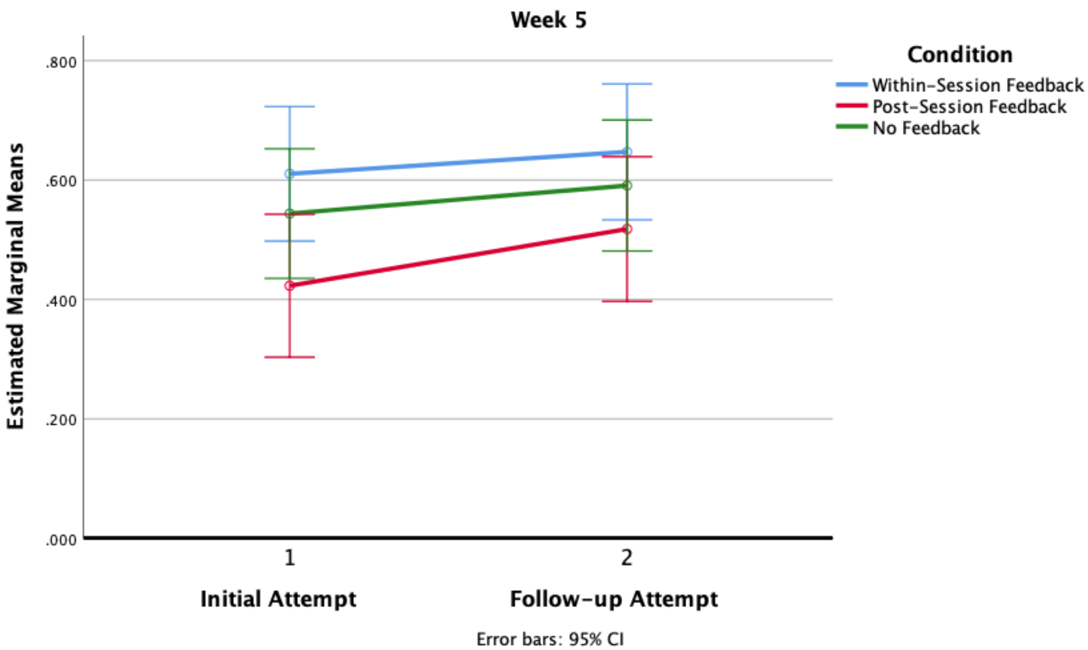


Week 5

A two-way repeated measures ANOVA revealed significant differences between initial pitch and rhythm scores and follow-up scores $F(1.000, 74.000) = 12.389$, $p = .001$, $\eta_p^2 = .143$. A Bonferroni adjusted pairwise comparison revealed no significant differences by condition.

Figure 4.14

Comparison of Week 5 initial attempt scores and follow-up scores by condition

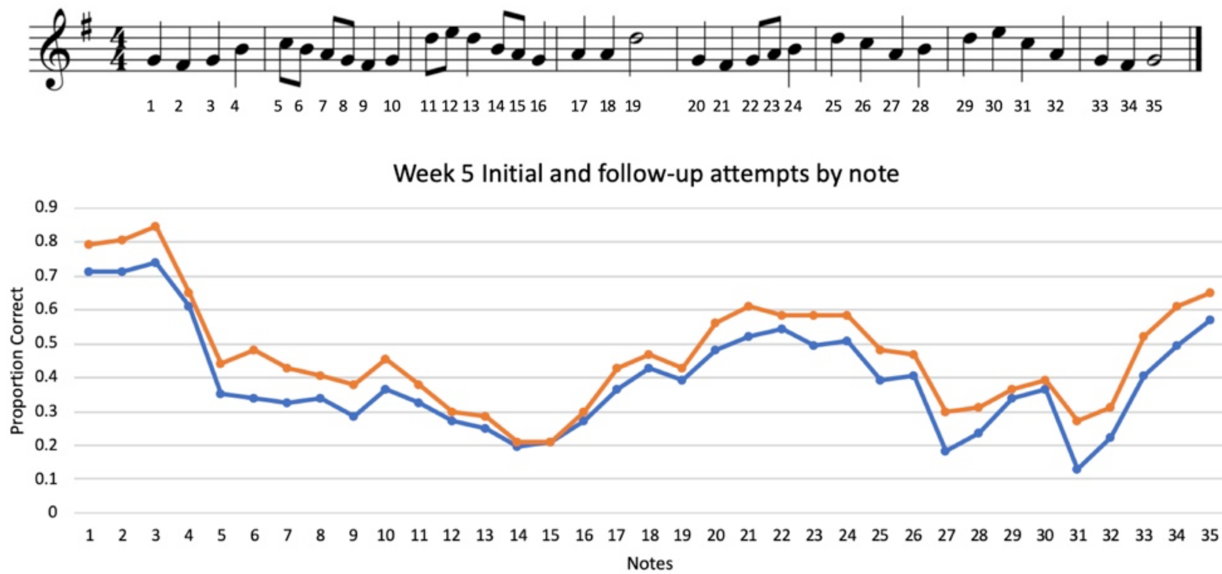


Note: Composite scores are reported as proportion of accurate pitches and rhythms

Like Week 4, participants in the PSF group posted the lowest initial composite scores and showed the greatest gains. Pitch and rhythm scores continued to diverge significantly during the initial attempt with rhythm scores averaging 14.2% higher $F(1.000, 74.000) = 32.886$ $p < .001$ $\eta_p^2 = .308$. Gains between pitch (.061) and rhythm (.058) were nearly identical. See Figure 4.15 for a display of the proportion correct by note during the initial and follow-up attempts.

Figure 4.15

Proportion Correct by Note During the Initial and Follow-up Attempts



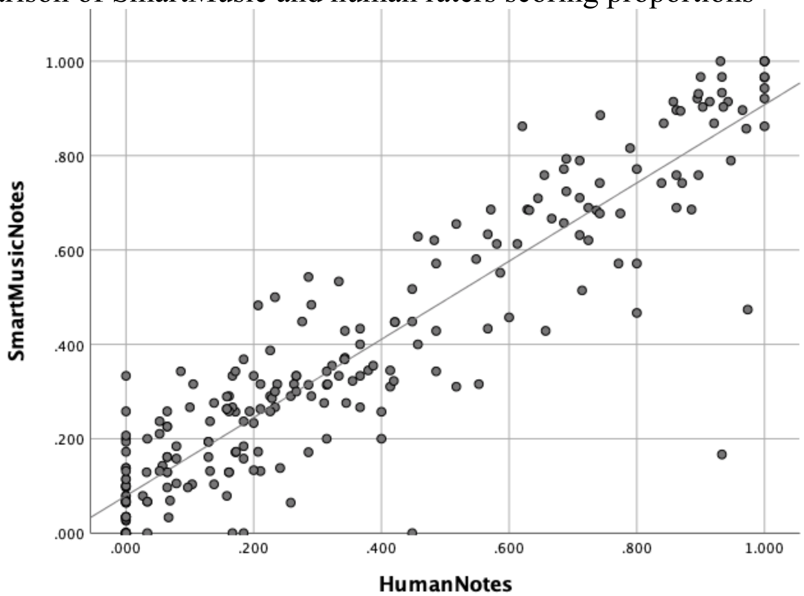
Research Question 3

I sought to examine the reliability of the feedback provided by the *SmartMusic* computer application when compared with human expert scoring. Visual feedback was presented once weekly to participants in the within-session feedback and post-session feedback groups. Approximately one-third (n = 237) of all attempts received feedback. Each note was considered accurate on the visual feedback when the notehead was colored green. I disregarded all other markings provided on the feedback. Notes were considered accurate by the human rater when both the pitch and rhythm were judged to be correct.

I analyzed reliability using a proportion of agreements divided by agreements plus disagreements (C. K. Madsen & C. H. Madsen, 1970). The proportion of agreement between *SmartMusic* and my scores (n=237) had a mean of 0.841 (SD = .124). Scores ranged between full agreement (1.0) and low agreement (0.167). The 95% confidence interval was between 0.825 and 0.857. Figure 4.17 provides a scatterplot of this relationship.

Figure 4.17

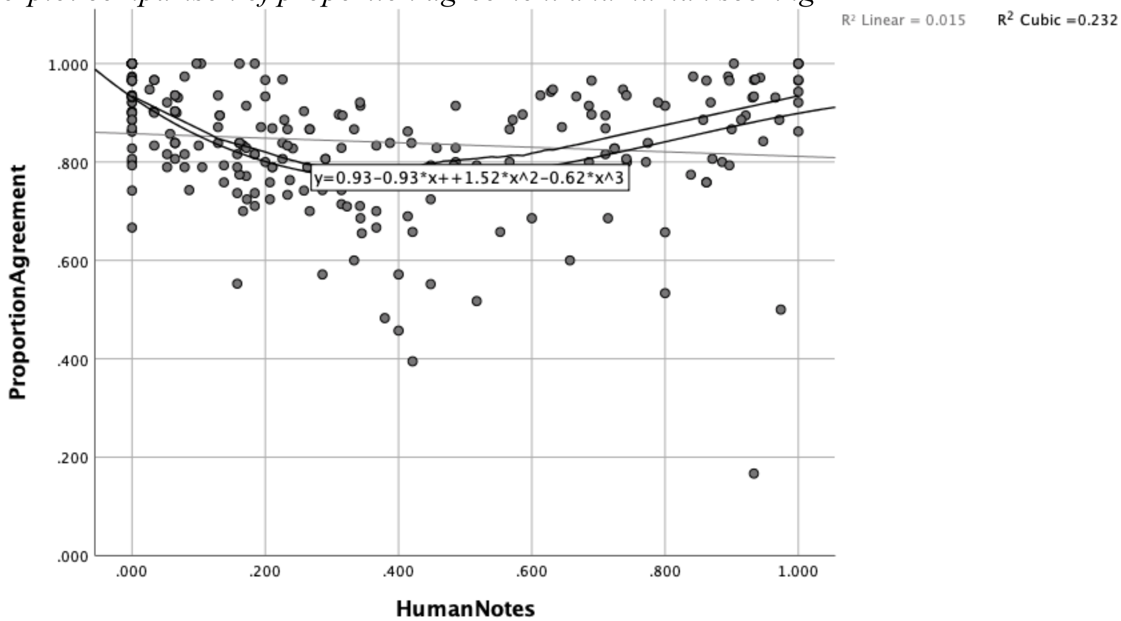
Scatterplot comparison of SmartMusic and human raters scoring proportions



To determine if a certain range of scores were more likely to relate with a positive relationship between SmartMusic and human scoring, I created a scatterplot with a fit-line contained in Figure 4.18. A U-shaped, cubic line was revealed, suggesting high scores and low scores were more likely to demonstrate agreement than scores in the middle two quartiles.

Figure 4.18

Scatterplot comparison of proportion agreement and human scoring



Conclusion

High school chorister volunteers (*n* = 77) from two suburban public schools completed a five-week sight-singing assessment session that was preceded by a pretest and followed by a posttest. Choristers, grades 9-12, included participants in auditioned (*n* = 48) and non-auditioned (*n* = 29) choral ensembles. More participants sang in the treble range (n = 51) when compared with those who sang the tenor/bass range (n = 26).

A comparison of pretest and posttest scores revealed a slight, non-significant improvement among participants from pretest to posttest. No significant differences were found by condition. Analysis of sight-singing scores on initial attempts during each weekly session revealed significant improvement between weeks one and three and weeks two, four, and five. A comparison of each initial attempt to follow-up attempts revealed significant gains for weeks one, two, four, and five. Participant gains during week three were not significant. A non-significant, negative relationship was found among singing scores with participants in the WSF group and the PSF group when compared to those in the NFC group. Voice part was not a significant predictor of accuracy, but school and choir placement were found to predict higher achievement. A comparison of *SmartMusic* feedback and human scoring revealed a very strong positive correlation when comparing proportion of correct notes $r(235) = .923$, $p < .001$. Analysis of agreements divided by possible agreements between human and SmartMusic revealed 84.1% consistency of scores.

In the next chapter, I will discuss the findings reported here as they relate to the research questions. I will also offer some possibilities as to why the feedback provided during this study did not have an effect on sight-singing achievement scores. I will conclude with a discussion of the implications of this study for music education.

Chapter 5

Discussion

The purpose of this study was to compare the effects of the presence and timing of computerized feedback provided by the *SmartMusic* interface on sight-singing achievement. Participants (*n* = 77) volunteered for a nine-week study that included a sight-singing pretest, a series of five weekly assessment sessions, and a sight-singing posttest. Each participant attempted seven different melodies a total of 12 times. Each melody singing attempt was scored for pitch and rhythm accuracy, and scores were converted into a proportion of correct pitches and rhythms.

Overall, the presence and timing of *SmartMusic* visual feedback did not affect student sight-singing achievement, either within-session, or pretest to posttest. The results suggest that students showed some improvement in their sight-singing scores, but that growth did not translate to the posttest task. The *SmartMusic* feedback showed potential for reliability when compared to the human rater but was not as reliable as an additional expert human rater.

This chapter will begin with an examination of participant data, attrition, and rater reliability. This will be followed by a discussion of the findings related to each research question. There will then be a discussion of the main findings of this study regarding the effects of feedback on sight-singing achievement. This chapter will conclude with a discussion of the limitations of this study and implications of these findings for classroom teachers and future technology development.

Participants

It should be assumed that the schools that participated in this study do not represent the 9-12th grade population of the United States. As such, the students who volunteered to participate in this study were not a representative sample of high school students in the United States. Because participant data was limited to choir enrollment and voice range, a demographic profile of race, socio-economic status, and English language learners could not be made. As with other sight-singing studies using volunteer participants, there existed a high likelihood of selection bias among participants. While I could not assess the sight-singing skills of those who did not volunteer, a distribution of scores from those who did tends to suggest that this study seemed to attract students who were either already fairly proficient and confident sight-singing, or by contrast, students who were unaware of how difficult it could be. The attrition rate for this study was low (4.8%), and as only two participants asked to withdraw from the study, it may be reasonable to assume that the procedures employed by this study, including group assignment, did not contribute to participant attrition.

Rater reliability

I scored all melody attempts (n = 924) and used an additional expert rater who scored approximately 28% (n = 265) of the sample selected randomly to determine rater reliability. A formula of agreements divided by agreements plus disagreements, yielded an agreement rate of .908 for pitch and an agreement rate of .852 for rhythm. These reliability numbers are lower than those provided by Henry (2011) who reported near 95% agreement. The use of the *SmartMusic* interface, including a visual quarter note indicator and a click track, required a more nuanced definition of accuracy when scoring. While our agreement for pitch scores was high, we did disagree almost 10% of the time. When this happened, it was primarily due to disagreements stemming from melody attempts that were near the 50-cent mark sharp or flat from the key. Differences were more pronounced when comparing rhythm scores. Consistently, when we differed, the additional expert rater was more lenient when scoring rhythm. I attribute these differences of interpretation regarding the third guideline under rhythm scoring, which stated, “These are high school students and you are trying to score their ability, not their precision. Some leeway is appropriate.” Though there were discrepancies in the validity of the rhythm accuracy construct, I felt the proportion correct score comparisons reported in the ICC were reliable enough to use my entire set of scores without modification.

Research Question 1

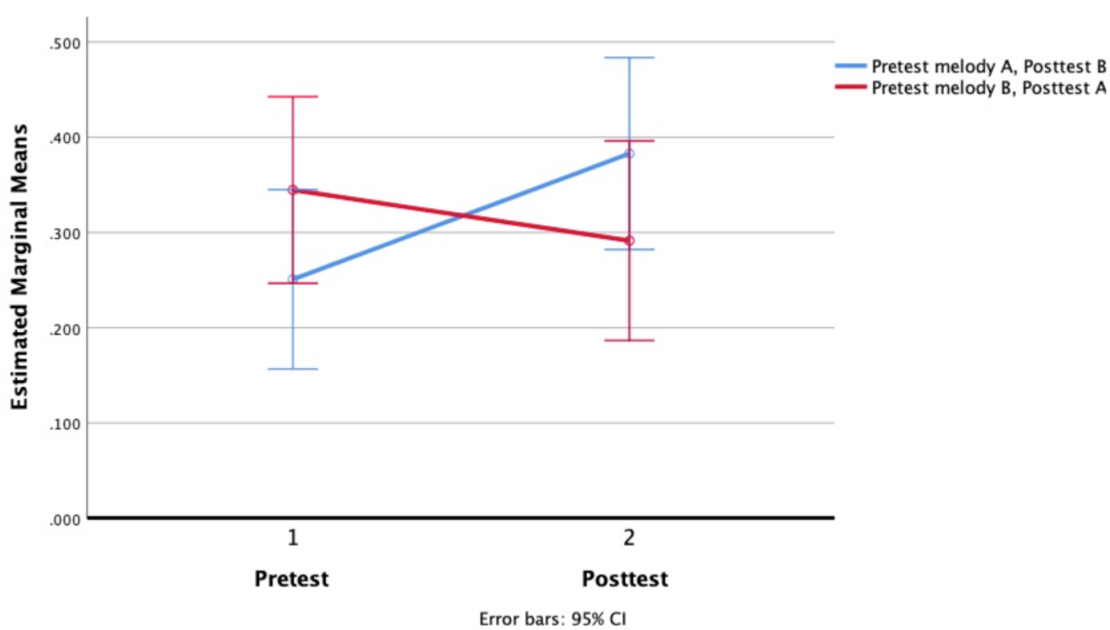
With the first research question, I sought to ascertain if the presence or timing of computerized feedback affected sight-singing achievement. Prior research has established the efficacy of individual assessment to improve sight-singing achievement (Demorest, 1998; Henry, 2014, Henry & Petty, 2014). However, under these conditions, significant improvements in sight-singing achievement were not found when comparing pretest and posttest scores. Furthermore, group assignment had no significant effect on sight-singing scores. Several factors may have contributed to this lack of improvement and are addressed below.

The pretest and posttest procedures were different than those experienced by the participants during the weekly sight-singing sessions. While the interface and initial practice time were identical, the click-track and quarter note indicator were disabled. It may be reasonable to assume that after 10 melody attempts with these features that some participants became accustomed to, if not reliant upon, these features when attempting a melody. It is unknown if participant scores would have improved if they had used the same interface they experienced during the weekly sessions. Evidence of individual improvement was found, however, by examining the progression of initial attempt scores achieved during each weekly session, though it is possible that this growth may be attributed to differences in melody. Figure 4.4 in Chapter Four, displays the significant improvement in initial composite scores from Week 1 to Week 5.

The research design may have been the cause of some issues. Melody A and Melody B differed more than I intended. Melody B was performed with significantly more accuracy than melody A on both the pretest and posttest, $F(1, 71) = 16.031$, $p < .001$, $\eta_p^2 = .184$. See Figure 5.1 for a graph of the differences. This disparity of difficulty made a comparison of gains difficult. As a result, it is possible that the differences in melody obscured possible gains made by participants.

Figure 5.1

Comparison of pretest and posttest scores by melody



The number of students who completed this study ($n = 77$), was just above the threshold required for making a six-group comparison between pretest and posttest scores. It is possible the model had insufficient power to find differences. Yet another possible issue was in the timing of the posttest. In order to get five weeks of uninterrupted assessment sessions, I had to place the posttest during the week that followed a district-wide, week-long vacation for the Thanksgiving holiday. It is possible that the week off negated gains that could have been measured by the posttest had the timing been different.

It should be noted that only participants in the Post-session Feedback (PSF) group showed negative gains from pretest to posttest. There may be a few possible explanations for this occurrence. First, participants in this group did not receive feedback until after a second attempt on each melody, and as a result, they may have learned to use the first attempt as practice during each weekly session, and therefore, were not prepared to use the 30-second practice time efficiently. Evidence of this potential may exist as the PSF group showed the lowest initial score during weeks 4 and 5 but also showed the greatest gains between the initial and follow-up attempts.

This study was not able to find significant differences between groups when comparing pretest and posttest scores. It remains unclear if either the frequency of sight-singing sessions, whether daily, weekly, or monthly, or duration of the treatment period, longer than five weeks, would have changed these findings. It should also be noted that though participants in the within-session feedback group scored the highest during the initial attempt each of the five weeks of the study, differences between groups maintained a parallel motion throughout the study suggesting that differences in scores were more likely the result of differences between groups that existed at assignment rather than result of research condition.

Question 2

Question two sought to ascertain if the presence or timing of feedback affected participants' ability to improve accuracy on the performance of a melody following an initial sight-singing attempt. This section will discuss the findings related to the differences between the initial and follow-up attempts during each week of the five-week treatment period.

The three-model regression analysis listed in Table 4.5, compared the scores of the follow-up attempt for every melody during the five-week treatment period to a series of predictor variables. Not surprisingly, the results revealed the greatest predictor of sight-singing achievement on the second attempt was the score of the initial attempt. This model revealed no significant differences by group assignment, suggesting feedback had no discernible effect in overall sight-singing achievement. Significant predictors were found among the choir enrollment with participants who were enrolled in more select choirs were more likely to have higher sight-singing scores on the follow-up attempt as listed in Table 5.1. This suggests that students who auditioned for, and were placed in more select choirs, were more likely to demonstrate sight-singing acumen, corroborating the findings of Demorest and May (1995).

Table 5.1

Average Initial Sight-singing Scores by Choir Enrollment

Choir	*n*=	Average initial composite score for weeks 1-5
Non-auditioned Choir	31	.242
Select Treble Choir	10	.501
Large Mixed Choir	18	.455
Small Mixed Choir	19	.685

Note: Composite score is reported as a proportion of correct pitches and rhythms

The regression model also showed School B was more likely to have higher scoring sight-singers than School A. This could be partially attributed to selection but may also be the result of sound contamination from the assessment room in school A, where the sound of the choir rehearsal can clearly be heard by the participants, possibly causing distraction.

An additional regression model was created to compare participants' gains made during all weeks of the five-week treatment period to determine if group assignment and other variables predicted differences. The first model compared group assignment and a small but significant negative relationship ($\beta = -.115$, $p < .05$) among participants in the WSF group when compared to the control group. Participants in the PSF group did not show any significant differences in gains when compared to the control. These findings reveal that students who were not given feedback prior to a second attempt showed significantly greater gains than those who received feedback, though it should be noted that participants in the within-feedback group scored consistently higher on the initial attempt, possibly limiting their potential growth when compared to the other groups.

The second regression model comparing gain scores added only one significant predictor of gains. Students in the Large Mixed Choir were more likely to show improvement than those in the non-auditioned ensembles. No other choir, including the more select Small Mixed Choir was found to predict gains, possibly suggesting that the melodies selected for this study were best suited to show improvement with students who were enrolled in the Large Mixed Choir. This may point to the benefits of teachers to selecting level-appropriate assessment samples.

The scatterplot (Figure 4.5) displaying the cubic relationship between average initial composite scores and average gains for all participants during weeks 1-5 further displays the need for level-appropriate melodies. Participants who averaged below 20% during their initial attempts were less likely to show as much improvement as those who scored between 20% and 80%, despite having the greatest opportunity for gains, corroborating the findings of Killian and Henry (2005). Those who averaged above 80% likely reached a ceiling effect as they approached 100% correct. Additionally, the feedback provided by *SmartMusic* was of no benefit to participants in the WSF group when compared with the other groups, even among those participants who scored below 20%.

Week by Week Analysis

A comparison of gains occurring between the initial and follow-up attempts reveals significant improvement for each group every week except Week 3. Differences among groups were non-significant and followed mostly parallel gains each of the five weeks of the study. This section will discuss the gains made within each session and includes a week-by-week comparison of some pitch and rhythm elements that were common among melodies.

Week 1. Among all weekly sessions, Week 1 presented the lowest mean score for both pitch (m = .320, SD = .181) and rhythm (M = .268, SD = .180) during the initial attempt but also showed the greatest gains for pitch (M = .112, SD = .050) and rhythm (M = .136, SD = .041). It may be reasonable to assume that some of the disparity between attempts one and two on this melody may be attributed to adapting to the *SmartMusic* interface features, including the quarter note indicator and click-track. It should be noted that the greatest gains in pitch, above 20% improvement, occurred when the accuracy of the first attempt was below 20% though several notes that initially scored below 20% did not show the same level of improvement.

It has been suggested that novice sight-singers tend to start at the beginning and run through a melody (Killian & Henry, 2005), possibly concentrating improvement in the early measures. However, gains between the first and second attempts were not located in any particular measure or area of the melody. During Week 1, scale degree did not seem to affect gains as all scale degrees improved between 9.2% for *ti* and 13.9% for *mi.*

Both Week 1 and Week 4 began with a similar *do re mi* pattern. Gains on these pitches were higher during Week 1 but by Week 4, participants averaged 83.1% correct on the first attempt. This seems to represent a plateau in participants' abilities to improve upon the first attempt as gains were negative for participants on these pitches and rhythms during Week 4. It should be noted that Henry (2013) found no differences between the keys of D-Major, E-flat-Major, and F-Major, but G-Major has not been studied. See Figure 5.2 for a comparison of the initial *do re mi* pitches from Week 1 and Week 4.

Table 5.2

Weeks 1 and 4. drm

Melody	pitch	*d*	*r*	*m*	gains	*d*	*r*	*m*
Week 1		.679	.641	.641		.090	.090	.090
Week 4		.831	.844	.844		-.062	-.075	-.049

Note: Week 1 Melody is G-Major and Week 4 melody is in D-Major

The dotted quarter note followed by an eighth-note rhythmic pattern that occurs in the first two counts of the third measure also appears in Week 4 in the first, sixth, and seventh measures, as listed in Table 5.3. Though improvements in this pattern were found between Week 1 and the initial attempt in Week 4, each subsequent attempt of that pattern during the Week 4 initial attempt shows a decline and gains during the second attempt of Week 4 were not found until final occurrence when the score dropped below that of Week 1. This likely corroborates the findings of Henry (2011), who found pitches were prioritized over rhythms.

Table 5.3

Weeks 1 and 4. Dotted-quarter note followed by an eighth note

Melody	rhythm *dotted-quarter*	*eighth*	gains *dotted-quarter*	*eighth*
Week 1 (third measure)	.295	.244	.128	.077
Week 4 (first measure)	.584	.364	-.065	.052
Week 4 (fifth measure)	.416	.299	-.026	.026
Week 4 (sixth measure)	.260	.182	.078	.104

Note: Though rhythmic patterns were identical, pitches and melodic context were not.

Week 2. Participants showed significant gains during Week 2 between the initial attempt and the follow-up attempt. Participants averaged positive gains on both pitch and rhythm except for notes three, four, and five of measure 5. Gains for pitch and rhythm were more modest than those found in Week 1, but the initial scores were significantly higher. Five pitches showed improvement greater than 10% and occurred in measures one, four, six, seven, and eight. Four of these pitches were within a step of tonic (*d, r,* and *m*) with the *s* in the final measure showing 11% improvement from the initial attempt.

Three melodies, those from Week 2, Week 3, and Week 5, began with identical melodic (*d t d m)* and rhythmic (quarter-notes) patterns. Improvement in the performance of these pitch skills can be seen when comparing Week 2 to Week 3 and Week 5. When participants in Week 3 averaged between 83.3% and 87.2% correct, however, they showed negative gains. The final note of this pattern shows improvement in the initial attempt each week and gains between the initial and follow-up attempts. These findings suggest that learning took place despite the presence of feedback having no apparent effect.

Table 5.4

Weeks 2, 3, and 5 dtdm pattern

Melody	pitch	*d*	*t*	*d*	*m*	gains	*d*	*t*	*d*	*m*
Week 2		.770	.714	.753	.532		.037	.102	.010	.046
Week 3		.872	.833	.872	.590		-.026	-.064	-.051	.026
Week 5		.792	.779	.779	.623		.052	.052	.078	.052

Note: Week 2 melody was in E-flat Major, Week 3 in F Major, and Week 5 in G Major

Week 3. The melody used for Week 3 was the only melody not to show significant gains between the initial and follow-up attempt. A clear cause of this lack of improvement was not readily found in the data. It is possible that the relatively high scores in the first measure represent a ceiling effect for these participants. Additionally, the melody contains more notes than any other week (n = 39), and participants may have been overwhelmed and not used their practice time efficiently. This Week 3 was also during concert week for both schools with participants in school A having just completed their concert, and those from School B were in their final on-stage rehearsals and used a different assessment room as a result.

More evidence of the possibility of a ceiling effect can be found in the final *do* of each weekly session melody. Week 3 scores were higher than any other week on the final note, and the final two notes were higher than the same notes in Week 5 as seen in Table 5.5. Gains for these notes, however where negative during Week 3 but positive for all others.

Table 5.5

Weeks 1, 2, 3, 4, and 5. final(t) d

Melody	pitch	*t*	*d*	gains	*t*	*d*
Week 1		-	.462		-	.077
Week 2		-	.584		-	.060
Week 3		.628	.692		-.064	-.038
Week 4		-	.506		-	.058
Week 5		.506	.597		.013	.065

Week 4. Participants in the post-session feedback group showed more growth during Week 4 than the other two groups. However, they did produce the lowest mean proportion correct on the initial attempt and may have been positioned to show more improvement. This trend where participants in the PSF group dip on the initial attempt but gain on the follow-up continues in Week 5. It is possible that those in the PSF groups began to view the initial attempt as a practice run while focusing their efforts on the follow-up attempt where they received feedback. Since the treatment period ended after five weeks, I cannot determine if this trend would continue.

Week 5. Scores during Week 5 were, on average, higher than any other week during both the initial and follow-up attempts, once again suggesting participants improved during the treatment period. The melody used during Week 5 had an identical pitch and rhythm pattern in the fifth measure that was initially seen in the fifth measure of the Week 2 melody (*d t, d r m*). While they showed very similar initial scores, the follow-up attempt showed greater gains during Week 5.

Table 5.6

Weeks 2 and 5. dtdrm

Melody	pitch	*d*	*t*	*d*	*r*	*m*	*gains*	*d*	*t*	*d*	*r*	*m*
Week 2		.584	.584	.610	.610	.571		.034	.021	-.005	-.031	.008
Week 5		.584	.584	.623	.558	.545		.078	.052	.039	.065	.078

Feedback

The main finding of this study is that the presentation of visual feedback provided by *SmartMusic* had no significant effect on sight-singing achievement. In this section, I will offer some suggestions as to why this may have been the case.

Sight-singing is a complex task. As a result, if one of the component skills of sight-singing is missing the singer is unlikely to be successful. As the ability to read, understand, and audiate written notation is a key sight-singing skill (Fournier et al., 2017; Vujović & Bogunović, 2012), it is logical to assume that if a student lacks a basic understanding of written notation, feedback using that notation is likely meaningless. Additionally, students who struggle to sing accurately, so called "poor-pitch singers," (Pfordresher & Brown, 2007) may understand the notation, but not sing with enough accuracy to produce a response from the visual feedback apparatus. Though not common, some participants in this study sang the correct solfege syllables in rhythm but failed to sing the pitches accurately.

Additionally, this study used a single, one-model approach as all participants sang the same melodies regardless of ability level and as a result, neglected the use of feedback to provide information informing future instruction (Fautley, 2010). Participants who struggled to maintain key, for example, received feedback that may not have presented useful information. See figure 5.2 for screenshots of the *SmartMusic* feedback provided participants that likely provided no useful information. A more effective use of this feedback may be to assign shorter melodies with a narrower range until the participant found some success. Additionally, participants who were able to perform the melody correctly on the first attempt were not able to make any improvements so the feedback, though confirming accuracy, did not provide assistance. A failure to use the feedback to alter the assessment may have narrowed its possible effectiveness to a small range of participants.

Figure 5.2

Excerpts from feedback provided to participants during Week 1

Sample from a participant who scored .00 for pitch and .00 for rhythm

Sample from participant who scored 1.00 for pitch and 1.00 for rhythm

Note: Scores were provided by a human rater

The timing of the feedback for this study, though varied by condition, was provided immediately following a melody singing attempt. Research has suggested that delayed feedback can reinforce learning and retention (Clariana, 2000; Nakata, 2015). It is possible that had the visual feedback been delayed 24 hours, it may have been more useful. Demorest (1998) utilized delayed feedback as participants in that study were given general comments and approximate scores after the attempts had been scored. The *SmartMusic* interface also includes a feature displaying the percentage correct for participants. This feature was covered up during this study. It is possible that presenting students with that information after a period of time could have enhanced learning. Additionally, this study's design did not allow for a fourth group, one where participants received feedback twice, once following the initial attempt and once after the follow-up attempt, or a fifth group where students were allowed to practice sight-singing without being recorded, thus eliminating the assessment piece.

Another possibility is that the feedback was too overwhelming and failed to present a clear path to improvement. Figure 5.3 displays feedback from a sight-singing attempt by a student who was very close to being accurate but was either late, in the wrong key, or a combination of both. The feedback failed to provide a clear description of what went wrong or how the melody attempt could be improved.

Figure 5.3

SmartMusic feedback may not offer much information

I made a technical support request to *MakeMusic*, the parent company that owns *SmartMusic* in March of 2020 asking them to define the tolerances for intonation allowed on the interface's assessment feature. Zachary C., responded by email to my inquiry, "Unfortunately, we don't have a breakdown available for *SmartMusic*'s assessment available for users" (email correspondence, 2020). The proprietary nature of the software and the unwillingness of *MakeMusic* personnel to reveal specific details about the technology, makes it less useful to students and teachers.

Question 3

With research question three, I addressed the reliability of the *SmartMusic* feedback compared to my scoring. Overall, agreements on the accuracy of notes where quite high considering the intricacies of the human voice. While I agreed with the additional human rater 91.5% of the time, I agreed with the *SmartMusic* apparatus 84.1% of the time. It would be reasonable, then for a choir teacher to assume that the feedback offers insight into choristers' sight-singing abilities. However, I would advise caution against using the assessment feature as a legitimate grading tool, especially in a high-stakes situation. I found the software to be susceptible to noise from the adjacent rehearsal space, and it struggled to assess quiet singers. On a few occasions, the feedback was simply wrong.

Anecdotally, I have heard choir directors share with other colleagues and me that the *SmartMusic* assessment feature encouraged their choristers to sing with an unnatural tone. After scoring 916 sight-singing attempts, I did not find this to be the case. Participants' vocal tone remained consistent throughout the treatment period and at no point did any student resort to a staccato singing style. I asked the additional rater if they noticed any change in vocal tone and they reported to me that they had not. It is possible that these habits, should they arise, would only do so after a longer exposure to *SmartMusic*. It should be noted that I did not analyze participants' singing using scientific acoustic measures including long-term average spectra measurements or fundamental frequency analysis (Grady, 2014).

Limitations

Due to the quasi-experimental nature of this study and limited scope of the participants, findings of this study cannot be generalized to the population as a whole. As with many sight-singing studies that ask for volunteers (Demorest, 2001), there is a high likelihood of selection bias among these participants. Findings are also specific to the procedures detailed in Chapter 3 and any change in those procedures would have the potential of producing different results. Findings of this study relating to feedback, are specific to the visual feedback provided by *SmartMusic*. It should not be assumed that because the feedback used in this study did not produce differences among groups that student access to feedback is not still an important part of the learning process. However, the limitations of this feedback offers teachers and researchers the opportunity to keep looking for effective ways of giving students useful information that leads to musical growth.

Future research

It is possible, that discernible differences would have been found among groups in this study if participants had been assigned sight-singing melodies that presented an appropriate level of challenge for their abilities. In such a situation, the feedback may have become more beneficial to a greater number of participants. Figure 4.5 in Chapter Four shows the apogee of the inverted-shaped U around .60 proportion correct, suggesting that about 60% correct was the level of challenge that allowed the greatest room for improvement following 90 seconds of practice. Any higher, and participants likely approached a ceiling effect, and as scores got lower, participants may have lacked the skills necessary to overcome the challenges presented by the melody (Killian & Henry, 2005). Sixty percent correct may be a good starting point for designing formative-type sight-singing assessments. Research suggests that self-efficacy and the belief that time-on-task will be productive play an important role in how research participants engage with feedback (Madsen & Duke, 1985; Timmers, et al., 2013). Future researchers could design and test sight-singing methods that track student self-efficacy for sight-singing and explore different feedback models that highlight improvement.

It was my hope that this study would help elucidate sight-singing instructional practices that help choristers gain independence. During this study, participants were only allowed to use the *SmartMusic* interface during weekly in-class assessment sessions. It is unknown if students given free access to the software would engage with it outside of rehearsal. It is possible that students who were motivated to learn sight-singing skills would practice on their own. Future researchers might examine how students self-regulate during sight-singing practice when using technology. Additionally, researchers have yet to quantify what level of sight-singing skill allows for chorister independence and under what conditions they are indeed independent.

This study is the first among the extant research literature I reviewed that allowed participants to attempt a melody again following an initial sight-singing attempt. It was encouraging that many students, regardless of feedback condition, were able to diagnose errors and correct them in a subsequent attempt. This study's design did not provide any insight into the processes with which the students undertook, either with the feedback or without, to go about correcting mistakes. Researchers could design a study where participants talk aloud while reflecting on a sight-singing attempt or while preparing a follow-up attempt. Eye-tracking technology may also offer insight into student interactions with feedback.

Implications

Even though not every participant in this study demonstrated improvement and the posttest failed to provide evidence of significant improvement, student scores found in the weekly sight-singing assessment sessions suggests that individual assessment did improve sight-singing achievement, corroborating earlier research (Demorest, 1998; Henry, 2015, Petty & Henry, 2014). This study provides evidence of the importance of adapting assessment difficulty to meet student abilities appropriately. The use of feedback should not be uni-directional as was the case in this study. Teachers who use technology like this should continually monitor and respond to student performance by altering instruction and future assessment. This technology may be best utilized as a supplement to good teaching but will not replace a quality teacher.

Students need to practice sight-singing skills in the same manner that they will be assessed. One possible reason for students' lack of improvement from the pretest to the posttest is that the posttest procedures did not match the assessment session procedures, or worded differently, the summative assessment procedures did not match the formative assessment practice. Students should not be expected to sight-sing individually when their only practice has been as a group. Furthermore, they should not be expected to perform without a metronome and quarter-note indicator if that was a regular part of their instruction.

Though the feedback provided by *SmartMusic* did not affect sight-singing achievement, the potential exists for teachers to use technology like *SmartMusic* to facilitate individual assessment and curate individual sight-singing attempts electronically. The need for teachers to be able to engage students in asynchronous instruction and assessment has become very pertinent (Chrysostomo & Triantafyllaki, 2020) so teachers should continue to explore the opportunities for students to engage with technology as a means of individual assessment.

The pitch recognition software used by *SmartMusic* likely demonstrated enough reliability that the potential for building an interactive platform exists. This technology might be useful to build a scaffolded interface that adjusts difficulty as participants improve using targeted pitch skills (Henry, 2004). The potential exists to create sight-singing software that is less academic in appearance and more engaging to the user. Using a video game model, software developers could build a program that balances challenge and skill to encourage a flow state (Chen, 2007; Cowley et al., 2008; Jin, 2012).

Conclusion

As long as choirs continue to perform music written in traditional Western notation, sight-singing will be an important skill in the development of choristers' musical independence. While the visual feedback used in this study did not yield any significant results, it did emerge as a viable tool to supplement educators' ability to teach sight-singing. The voice-pitch recognition software used for this study has potential benefits, but it is incumbent upon software developers and teachers to use it in a manner that promotes student learning. As we continue to make individual assessment more efficient and effective, teachers will be able to give choristers the best tools possible to make music independently and enjoy a lifetime of reading choral music.

References

American Choral Directors Association https://acda.org/ACDA/Repertoire_and_Resources/Youth/Children_and_Community_Youth/Standards.aspx

Armstrong, M. (2001). Adjudicated sight-reading for the choral ensemble: An incentive for music literacy. *Choral Journal, 41*(10), 21-30.

Atlas, A. W. (1998). *Renaissance music: music in Western Europe, 1400-1600*. WW Norton.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of educational research*, *61*, 213-238.

Banton, L. J. (1995). The role of visual and auditory feedback during the sight-reading of music. *Psychology of music*, *23*(1), 3-16.

Bennett, P. (1984). Tricks, masks, and experience: Is imitation passing for music reading? *Music Educators Journal, 71*(3), 62-69.

Benton, C. W. (2002). *A study of the effects of metacognition on sight-singing achievement and attitudes among students in a middle school choral music program* [Doctoral dissertation, Shenandoah University].

Boisen, R. (1981). The effect of melodic context on student's aural perception of rhythm. *Journal of Research in Music Education, 29*(3), 165-172.

Brendell, J. K. (1996). Time use, rehearsal activity, and student off-task behavior during the initial minutes of high school choral rehearsals. *Journal of Research in Music Education, 44*, 6-14.

Brown, K. D. (2001). *Effects of fixed and movable sightsinging systems on undergraduate music* *students' ability to perform diatonic, modulatory, chromatic, and atonal melodic passages* [Doctoral dissertation, University of Oregon].

Buck, M. W. (2008). *The efficacy of SmartMusic® assessment as a teaching and learning tool* [Doctoral dissertation, University of Southern Mississippi].

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281.

Butzlaff, R. (2000). Can music be used to teach reading? *Journal of Aesthetic Education, 34*(3), 167-178.

Cassidy, J. W. (1993). Effects of various sightsinging strategies on nonmusic majors' pitch accuracy. *Journal of Research in Music Education, 41*(4), 293-302.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31-34.

Chrysostomou, S. & Triantafyllaki, A. (2020) Transitioning to Online Music Teacher Education: Challenges and Opportunities for Knowledge Development in Ferdig, R. E., Baumgartner, E., Hartshorne, R., Kaplan-Rakowski, R., & Mouza, C. (eds). Teaching, technology, and teacher education during the covid-19 pandemic: Stories from the field. *Waynesville, NC, USA: Association for the Advancement of Computing in Education (AACE)*.

Clariana, R. B. (2000). Feedback in computer-assisted learning. *NETg University of Limerick Lecture Series*.

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

Colwell, R. (1963). An investigation of musical achievement among vocal students, vocal-instrumental students, and instrumental students. *Journal of Research in Music Education, 11*, 123-130.

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment*, *6*(2), 1-27.

Daniels, R. D. (1986). Relationships among selected factors and the sight-reading ability of high school mixed choirs. *Journal of Research in Music Education, 34*, 279-289.

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

Demorest, S. M. (2001). *Building choral excellence: Teaching sight-singing in the choral rehearsal*. Oxford University Press.

Demorest, S. M. (2004). Choral sight-singing practices: revisiting a web-based survey. *International Journal of Research in Choral Singing, 2*(1), 3-10.

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

Dempsey, J. V. (1993). *Interactive instruction and feedback*. Educational Technology.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dwiggins, R. (1984). Teaching sight-reading in the high school chorus. *Update: Applications of Research in Music Education, 6*(2), 8-11.

Farenga, J. (2013). *Arizona high school choral educators' attitudes toward the teaching of group* *sight singing and preferences for instructional practices*. [Doctoral dissertation, Arizona State University].

Fautley, M. (2010). *Assessment in music education*. Oxford University Press.

Fisher, R. A., Summitt, N. L., & Jacks, L. (2015). A description of sight-singing practices in Arkansas secondary choral programs. *Missouri Journal of Research in Music Education*, (52) 19-29.

Owen, S. L. (2015). *Student perceptions of the efficacy of SmartMusic practice software*. [Doctoral dissertation, California State University].

Floyd, E. & Bradley, K. D. (2006). Teaching strategies related to successful sightsinging in Kentucky choral ensembles. *Update: Applications of Research in Music Education, 25*(1), 70-81.

Fournier, G., Moreno Sala, M. T., Dubé, F., & O'Neill, S. (2017). Cognitive strategies in sight-singing: The development of an inventory for aural skills pedagogy. *Psychology of Music*, 0305735617745149.

Frey-Clark, M. (2017). Pitch Systems and Curwen Hand Signs: A Review of Literature. *Update: Applications of Research in Music Education*, *36*(1), 59-65.

Furby, V. J. (2008). *Process and product: The sight singing backgrounds and behaviors of first year undergraduate students* [Doctoral dissertation, The Ohio State University].

Gagné, R. M., Wager, W., & Rojas, A. (1981). Planning and authoring computer-assisted instruction lessons. *Educational technology*, *21*(9), 17-26.

Goolsby, T. W. (1994a). Eye movement in music reading: Effects of reading ability, notational complexity, and encounters. *Music Perception, 12*(1), 77-96.

Goolsby, T. W. (1994b). Profiles of processing: Eye movements during sightreading. *Music Perception*, 12(1), 97-123.

Goss, D. A. (2010). *Sight-singing assessment: A study of current beliefs and practices of Georgia middle and high school choral directors* [Doctoral dissertation, Capella University].

Grady, M. L. (2014). Effects of traditional pattern, lateral-only, and vertical-only conducting gestures on acoustic and perceptual measures of choir sound: An exploratory study. *International Journal of Research in Choral Singing*, *5*(1), 39-59.

Gurley, R. (2012). *Student perception of the effectiveness of SmartMusic as a practice and assessment tool on middle school and high school band students* [Doctoral dissertation, Texas Tech University].

Harrison, C. S. (1990). Relationships between graded in the components of freshman music theory and selected background variables. *Journal of Research in Music Education, 38*(3), 175-186.

Harrison, C. S., Asmus, E. P., & Serpe, R. T. (1994). Effects of musical aptitude, academic ability, music experience, and motivation on aural skills. *Journal of Research in Music Education*, 42(2), 131-144.

Hargiss, G. (1962). The acquisition of sight singing ability in piano classes for students preparing to be elementary teachers. *Journal of Research in Music Education*, 10(1), 69-75.

Hawkins, J. A. (2018). *Secondary choral music educators' use of technology-assisted assessment tools* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Henry, M. (2015). Vocal sight-reading assessment: Technological advances, student perceptions, and instructional implications. *Update: Applications of Research in Music Education*, *33*(2), 58-64.

Henry, M. L. (2013). The Effect of Key on Vocal Sight-Reading Achievement. *Texas Music Education Research*, *3*, 8.

Henry, M. L. (2011). The effect of pitch and rhythm difficulty on vocal sight-reading performance. *Journal of Research in Music Education*, *59*(1), 72-84.

Henry, M. (2004). The use of targeted pitch skills for sight-singing instruction in the choral rehearsal. *Journal of Research in Music Education, 52*(3), 206-217.

Henry, M. L., & Demorest, S. M. (1994). Individual sight-singing achievement in successful choral ensembles. *Update: Applications of Research in Music Education, 13*(1), 4-8.

Hime, L. J., Miksza, P., & Hunsucker, A. (2014). Admissions criteria and upper-divisional ratings as predictors of student teaching achievement in music education. *Bulletin of the Council for Research in Music Education*, (200), 7-22.

Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. *Interactive instruction and feedback*, 105-132.

Howard, D. M. (2005). Technology for real-time visual feedback in singing lessons. *Research Studies in Music Education*, *24*(1), 40-57.

Ittzés, M. (2004). Zoltán Kodály 1882-1967: Honorary President of ISME 1964-1967.

Jacobssen O. I. (1942) An analytical study of eye-movements in reading vocal and instrumental music. *The Journal of Musicology, 3*(4), 197-226.

Jeffries, T. B. (1967). The effects of order of presentation and knowledge of results on the aural recognition of melodic intervals. *Journal of Research in Music Education*, *15*, 179-190.

Jin, S. A. A. (2012). “Toward integrative models of flow”: Effects of performance, skill, challenge, playfulness, and presence on flow in video games. *Journal of Broadcasting & Electronic Media*, *56*(2), 169-186.

Jonassen, D. H. (1990). Thinking technology: Toward a constructivist view of instructional design. *Educational technology*, *30*(9), 32-34.

Karagiorgi, Y., & Symeou, L. (2005). Translating constructivism into instructional design: Potential and limitations. *Journal of Educational Technology & Society*, *8*(1), 17-27.

Karas, J. B. (2005). *The effect of aural and improvisatory instruction on fifth-grade band students' sight reading ability*. [Doctoral dissertation, The University of Nebraska-Lincoln].

Keene, J. A. (2009). *A history of music education in the United States*. Glenbridge Publishing Ltd..

Kegerreis, R. I. (1970). History of the high school a cappella choir. *Journal of Research in Music Education*, *18*(4), 319-329.

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

Killian, J. N. & Henry, M. (2005). A comparison of successful and unsuccessful strategies in individual sight singing preparation and performance. *Journal of Research in Music Education*, *53*, 51-65.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

Knox, M. (2003). *Reading music and written text: The process of sight- singing* [Doctoral dissertation, The University of Arizona].

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

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

Kuehne, J. M. (2007). A survey of sight-singing instructional practices in Florida middle-school choral programs. *Journal of Research in Music Education, 55*, 115-128.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*(2), 211-232.

Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. *Interactive instruction and feedback*, *320*.

Larson, R. L. (1977). Relationships between melodic error detection, melodic dictation, and melodic sightsinging. *Journal of Research in Music Education, 25*, 264-271.

Leong, S., & Cheng, L. (2014). Effects of real-time visual feedback on pre-service teachers' singing. *Journal of Computer Assisted Learning*, *30*(3), 285-296.

Lorek, M. J. (1991). Computer analysis of vocal input: A program that simulates college faculty sight singing evaluation. *Computers in Music Research, 3*, 121-138.

Lovorn, T. (2016). The Effect of Writing Solfège Syllables into Choral Repertoire on the Sight-Reading Ability of High School Choir Students. *Texas Music Education Research*, *15*, 24-31.

Lucas, K. V. (1994). Contextual condition and sight-singing achievement of middle school choral students. *Journal of Research in Music Education*, 42, 203-216.

Madsen, C. K. & Duke, R. A. (1985). Perception of approval/disapproval in music. *Bulletin of the Council for Research in Music Education, 85,* 119-130.

Madsen, C. K., & Madsen, C. H. (1970). *Experimental research in music*. Prentice Hall.

MakeMusic(https://www.smartmusic.com/blog/preparing-to-teach-music-remotely-utilizing-smartmusic-and-other-resources/, retrieved, March, 31st, 2020

Mark, M. L. (1994). *Source readings in music education history.* Charlottesville, VA: Lincoln-Rembrandt Publishing Inc.

Mark, M. L., & Gary, C. L. (1992). *A history of American music education*. Schirmer Books.

McClung, A. C. (2008). Sight-singing scores of high school choristers with extensive training in movable solfège syllables and Curwen hand signs. *Journal of Research in Music Education*, *56*, 255-266.

McClung, A. C. (2001). Sight-singing systems: Current practice and survey of allstate choristers. *Update: Applications of Research in Music Education, 20*(1), 3-8.

McGill, S., & Stevens Jr., H. M. (2003). *90 days to sight reading success: A singer's resource for competitive sightsinging.* AMC Publications.

McNaught, W. G. (1899-1900). The psychology of sight-singing. Proceedings of the Music Association, 26th Sess., 35-55.

Middleton, J. A. (1984). Develop choral reading skills. Music Educator's Journal, 70(7), 29-32.

Mory, E. H. (2004). Feedback research revisited. *Handbook of research on educational communications and technology*, *2*, 745-783.

Mosher, R. M. & Clement, B. B. (1925). Study of the group method of measurement of sightsinging. *Music Supervisors' Journal, 12*(2), 22-28.

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, *3*, 222-229.

Music Educators National Conference (1994) https://web.archive.org/web/20080424085956/http://www.menc.org/resources/view/national-standards-for-music-education

Myers, G. C. (2008). *Sight-singing instruction in the undergraduate choral ensembles of colleges and universities in the Southern Division of the American Choral Directors Association: Teacher preparation, pedagogical practices and assessed results* [Doctoral dissertation, The University of North Carolina at Greensboro].

Nakata, T. (2015). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*, *19*(4), 416-434.

National Association for Music Education (2014) https://nafme.org/wp-content/files/2014/06/5-Core-Music-Standards-Ensemble-Strand1.pdf

Nichols, B. (2012). The way we do the things we do: A survey of middle-school choral educators' sight-singing attitudes and influences. *Contributions to Music Education*, 87-100.

Nielsen, L. D. (2013). Developing musical creativity: Student and teacher perceptions of a high school music technology curriculum. *Update: Applications of Research in Music Education*, *31*(2), 54-62.

Nolker, D. B. (2006). The relationship between large ensemble sight-reading rating and the individuals' sight-singing success. *Missouri Journal of Research in Music Education, 43*, 3-16.

Norris, C. E. (2003). The relationship between sight singing achievement and melodic dictation achievement. *Contributions to Music Education*, 30(1), 39-53.

Norris, C. E. (2004). A nationwide overview of sight-singing requirements of large group choral festivals. *Journal of Research in Music Education, 52*(1), 16-28.

Owen, S. L. (2015). *Student perceptions of the efficacy of SmartMusic practice software* [Doctoral dissertation, California State University, Long Beach].

Ozeas, N. L. (1991). *The effect of the use of a computer assisted drill program on the aural skill development of students in beginning Solfege (interval identification and sight singing*) [Doctoral dissertation, University of Pittsburgh].

Paney, A. S., & Tharp, K. L. (2019). The effect of concurrent visual feedback on adult singing accuracy. *Psychology of Music*, 0305735619854534.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (eds.), (2003). *Knowing what students know: The science and design of educational assessment.* National Academy Press.

Perry, P. J. (2014). *The effect of flexible practice computer-assisted instruction and cognitive style on the development of music performance skills of high school instrumental students* [Doctoral dissertation, Shenandoah University].

Petty, C., & Henry, M. L. (2014). The Effects of Technology on the Sight-Reading Achievement of Beginning Choir Students. *Texas Music Education Research*, *23*, 28.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness."

Music Perception, 25(2), 95-115.

Platte, J. D. (1981). *The effects of a microcomputer-assisted instructional program on the ability of college choral ensemble members to sing melodic configurations at sight* [Doctoral dissertation, Ball State University].

Potter, A. A. (2015). *Sight-singing systems in collegiate choral curricula: An examination of conductors' best practices at degree-granting institutions of the National Association of Schools of Music* [Doctoral dissertation, The Florida State University].

Price, H. E. (1992). Sequential patterns of music instruction and learning to use them. *Journal of Research in Music Education*, *40*, 14-29.

Rutkowski, J., & Miller, M. S. (2003). The effect of teacher feedback and modeling on first graders' use of singing voice and developmental music aptitude. *Bulletin of the Council for Research in Music Education, 156*, 1-10.

Schmidt, C. P. (1995). Attributions of success, grade level, and gender as factors in choral students' perceptions of teacher feedback. *Journal of Research in Music Education*, *43*, 313-329.

Sight Reading Factory and https://www.sightreadingfactory.com, March 31st, 2020

Skinner, B. F. (1965). *Science and human behavior*. Simon and Schuster.

Snider, M. T. (2007). *Choral sightsinging strategies in 1A, 3A, 4A, and 6A Kansas high schools: Instructional time allotted, teacher strategies, and materials used.* [Masters thesis, Emporia State University].

Steinhauer, K., & Grayhack, J. P. (2000). The role of knowledge of results in performance and learning of a voice motor task. *Journal of Voice*, *14*(2), 137-145.

Timmers, C. F., Braber-Van Den Broek, J., & Van Den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & education*, *60*(1), 25-31.

Timoshenko, M. (2018). Seeing into the music score: eye-tracking and sight-reading in a choral context. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (p. 77). ACM.

Thorndike, E. L. (1933). A proof of the law of effect. *Science, 77*, 173-175.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology, 39*, 212-222.

Thostenson, M. S. (1967). The study of certain problems in eartraining related to achievement in sight-singing and music dictation. *Bulletin of the Council for Research in Music Education, 11*(1), 14-35.

Tucker, D. W. (1969). *Factors related to musical reading ability of senior high school students participating in choral groups* [Doctoral dissertation, University of California – Berkeley].

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research*, *85*(4), 475-511.

Von Kampen, K. E. (2003). *An examination of factors influencing Nebraska high school choral directors' decisions to use sight-singing instruction* [Doctoral dissertation, The University of Nebraska – Lincoln].

Vispoel, W. P., & Austin, J. R. (1993). Constructive response to failure in music: The role of attribution feedback and classroom goal structure. *British Journal of Educational Psychology*, *63*(1), 110-129.

Vujović, I., & Bogunović, B. (2012, July). Cognitive strategies in sight-singing. In *Proceedings Book of 12th International Conference on Music Perception and Cognition (ICMPC) and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music,* Thessaloniki, Greece.

Walls, K. C., Erwin, P. M., & Kuehne, J. M. (2013). Maintaining efficient ensemble rehearsals without sacrificing individual assessment: *SmartMusic* assessment could leave the director on the podium. *Journal of Technology in Music Learning*, *5*(1), 4-16.

Welch, G. F. (1985). Variability of practice and knowledge of results as factors in learning to sing in tune. *Bulletin of the Council for Research in Music Education*, 238-247.

Wilson, P. H., Lee, K., Callaghan, J., & Thorpe, C. W. (2008). Learning to sing in tune: Does real-time visual feedback help? *Journal of interdisciplinary music studies*, *2*.

Appendixes

Appendix A, Sight-singing Scoring

For the purposes of this study, every pitch and every rhythm will be scored separately. Therefore, every note will be worth two points.

Notes/Pitches

1. The melody be sung in the original key.
2. Participants may choose which octave they prefer.
3. Pitches may be within 50 cents of a half-step in either direction to be considered accurate.
4. Use of a solmization system is not scored- only pitch accuracy.
5. Pitches do not need to align with the click track to be considered accurate.
6. Participants may correct a single pitch by changing notes, sliding, or scooping, but may not go back once a subsequent pitch is attempted.
7. You may use clues to ascertain the participant's intended pitch including the click track, prior and following notes, and solfege syllables.
8. Accurate pitches are given a 1 (per note).
9. Inaccurate or omitted pitches are given a 0.

Rhythms

1. Rhythms must align with the click track and the quarter note indicator.
2. The pitch of a given note does not need to be accurate for the rhythm to be considered correct.
3. These are high school students and you are trying to score their ability, not their precision. Some leeway is appropriate.
4. Notes do not have to be performed for their full duration, but another note cannot be started before the current note duration is competed. Note pairs (two eighth-notes or a dotted quarter -eighth note pair) will likely need to be performed correctly to mark either correct in most situations.
5. The rhythm is considered incorrect if they change pitch or syllable during the note
6. Accurate rhythms are given a 1 (per note)
7. Inaccurate or omitted rhythms are given a 0

Other comments

1. In order for both pitches and rhythms to both be scored as accurate, the note and rhythm must be accurate.
2. When there is a discrepancy between pitches and rhythms, give preference to scoring the pitch as correct and mark the rhythm wrong.

Appendix B, IRB Approval

Northwestern | RESEARCH

Northwestern University
Institutional Review Board
Biomedical IRB
750 N. Lake Shore Dr., 7th Fl.
Chicago, Illinois 60611

irb@northwestern.edu
Office 312. 503. 9338

Social & Behavorial Sciences IRB
600 Foster St., 2nd Floor
Evanston, Illinois 60208

sbsirb@northwestern.edu
Office 847. 467. 1723

APPROVAL OF NEW STUDY

DATE: September 24, 2019

TO: Dr. Sarah Bartolome
FROM: Office of the IRB

DETERMINATION DATE: 9/24/2019
APPROVAL DATE: 9/20/2019

The Northwestern University IRB reviewed and approved the submission described below:

Type of Submission:	Initial Study
Review Level:	Expedited
Expedited Category:	- (6) Voice, video, digital, or image recordings - (7) Behavioral research/social science methods
Title of Study:	The effects of feedback on sight-singing achievement
Principal Investigator:	Sarah Bartolome
IRB ID:	STU00210808
Funding Source:	Name: The Graduate School
IND, IDE, or HDE:	None
Documents Reviewed:	• Sight-singing feedback recruitment, Category: Recruitment Materials; • Sight-singing feedback, Category: IRB Protocol; • Sight-singing feedback permission/assent, Category: Consent Form; • Effects of feedback debrief, Category: Debriefing Script;
Special Determination(s):	Children;

In conducting this study, you are required to follow the requirements listed in the Northwestern University (NU) Investigator Manual (HRP-103), which can be found by navigating to the policy section of the IRB website. Additionally, as Principal Investigator (PI) of this research study, you are expected to adhere to the investigator responsibilities outlined in the "What are my obligations as Investigator in order to conduct Human Research" section of the Investigator Manual (HRP-103).

If your study is a clinical trial, there are additional requirements including trial registration and results reporting on ClinicalTrials.gov. Federally-funded clinical trials are also required to post one IRB approved consent form, used during enrollment, on a publicly available federal website such as ClinicalTrials.gov. Please visit the clinical trials page on the IRB website for more information. If you would like an account created or need other assistance with ClinicalTrials.gov, please email clinicaltrials.gov@northwestern.edu.

An annual continuing review is not required for this project. The study team must still submit: modifications for project changes; RNIs (reportable new information); and a Continuing Review to close the project when it ends (for guidance on when a project can be closed, see GUIDANCE on Study Closure – HRP-1901.

NU IRB approval does not constitute or guarantee institutional approval and/or support. Investigators and study team members must comply with all applicable federal, state, and local laws, as well as NU Policies and Procedures, which may include obtaining approval for your research activities from other individuals or entities.

For IRB-related questions, please consult the NU IRB website at http://irb.northwestern.edu. For general research questions, please consult the NU Office for Research website at www.research.northwestern.edu.

Additionally, please note that the analyst who you worked with during the initial review and approval of your study is not the analyst that is responsible for the review of any subsequent modifications, continuing reviews, or RNIs. As such, please direct any further questions about modifications, continuing reviews, or RNIs to the analyst assigned to the subsequent submission.

Appendix C, School District Permission

[redacted]

Project Screening Action – District Level

To: Adam White

From: [redacted]
Assessment & Research

Date: 9/26/3019

Project Title: The effects of feedback on sight-singing achievement

Your research project has been reviewed and the project has been:

x approved

not approved

[redacted]

conditional approved based in changes to be made

Clarification/Comments:

Permission to proceed with working with [redacted] administrative teams at [redacted] to work out the logistical details.

This project has been assigned the following number for identification purposes:

Project Number: 2020_AW_17

Please submit a copy of the completed project to our office.

If further clarification is needed concerning this action, please contact:

[redacted]

Appendix D, Building Administrator Permission

School A

From: [redacted]
Date: Friday, September 27, 2019 at 9:15 AM
To: Adam White <adamwhite2020@u.northwestern.edu>
Cc: [redacted], Sarah J Bartolome <sarah.bartolome@northwestern.edu>
Subject: RE: Research study with [redacted]

Thank you Adam. [redacted] are on-board, I am too.

I appreciate the communication.

[redacted]

School B

[redacted]

September 30, 2019 at 8:48 AM

Details

LG

Re: Research study with [redacted]

To: Adam White, Cc: [redacted] Sarah J Bartolome

Good morning,

Yes, I approve this study at [redacted] I appreciate your thorough description and documents provided. I wish you the best as you complete the dissertation study.

Thank you,
[redacted]

Appendix E, Parent Consent, Student Assent form

Dear Parents and Guardians of Choral Students at [redacted] High School:

A researcher from Northwestern University is conducting a study on sight-singing achievement among high school chorus students entitled, "The effects of feedback on sight-singing achievement," IRB ID STU00210808. Your school, [redacted], has agreed to allow that researcher to assess choral students' sight-singing abilities in hopes of improving sight-singing instruction.

All sight-singing sessions will occur during regularly schedules choir rehearsals and will amount to approximately five minutes of time a week for the first nine weeks of the fall semester. Audio and screen-capture video recordings will be made in each session. At no time will a student's image or video be captured.

Students are not required to participate in this study, and nothing will be held against them if they choose not to do so. Prior to participating in the study, parents and students will review the attached permission and assent form and indicate they are willing to participate. If they are willing to participate, parents and students will sign the form and return it to your choir teacher. If they choose not to participate, the form will not be returned and students will participate in an alternate sight-singing session selected by the choir director.

If you have questions, concerns, or complaints, about the research, you can contact the research team at sarah.bartolome@northwestern.edu. This research has been reviewed and approved by an Institutional Review Board ("IRB"). You may talk to them at (312) 503-9338 or irb@northwestern.edu if your questions, concerns, or complaints are not being answered by the research team; you want to talk to someone besides the research team; or you have questions about the rights of research participants.

Best regards,
Adam G. White
PhD Candidate, Music Education – Northwestern University
adamwhite2020@u.northwestern.edu
(785) 845-5520

Dr. Sarah Bartolome – Principle Investigator
Professor of Music Education – Northwestern University
Sarah.bartolome@northwestern.edu
(847) 476-1682

Title of Research Study: The Effects of Feedback on Sight-singing Achievement
Principal Investigator: Sarah Bartolome, PhD, Northwestern University
IRB ID: STU00210808

Key information about this research study:

The following is a short summary of this study to help you decide whether to permit your child to be a part of this study.
The purpose of this study is to examine the effectiveness of technology as a sight-singing assessment tool. Over the period of nine weeks, your child will be asked to sight-sing a total of seven times with each session taking less than 5 minutes. All sessions will be completed in school during choir. There are no risks involved with this research though students may experience nervousness while engaging in sight-singing activities.

Why am I being asked to take part in this research study?

You are being asked to participate because you are a high school chorister in a program that teaches sight-reading skills and values musical independence.

What should I know about this research study?

In this study, we want to find out more about how computerized sight-singing programs can improve individual sight-singing achievement. You do not have to be in this study if you do not want to do so. If you do not want to participate, please select that option below and your teacher will give you an alternative sight-singing assessment. Your decision will not be held against you. All participant information will be anonymous and sight-singing scores will remain confidential.

What happens if I say "Yes, I want to be in this research?"

If you and your child decide to participate in this study, your child will be issued a participation code to ensure anonymity during data collection. For scientific reasons, this consent information does not include a complete description of the conditions being tested. When the study is completed, more information will be provided about the study and student participants will have the opportunity to ask questions.

Following confirmation of participation, student participants will sight-sing a short passage displayed on an iPad following 30 seconds on non-directed practice. This baseline sight-singing assessment will be captured using an audio recording device and a small video camera directed at the iPad. These recordings will be scored, and all student participants will be assigned to one of three matched groups that will be revealed at the completion of data collection.

Approximately one week after the baseline sight-singing assessment, student participants will engage in five-minute sight-singing assessment sessions, occurring once a week for five weeks. Each session will entail the same procedures. Student participants will enter the assessment area and identify themselves by their participation code. The researcher will begin the audio and screen-video recordings and will say the participant code aloud. The participant will then be presented with a sight-singing passage using the SmartMusic interface on an iPad. The researcher will leave the room and close the door. They will have 30 seconds to practice the passage and will make an initial attempt. Following the initial attempt, the participant will have 60 seconds to practice. Following the practice period. The researcher will enter the assessment area and restart the SmartMusic interface and exit the room. Following an additional 30 seconds of practice time, the student participant will attempt the sight-singing passage a second time. At the conclusion of the second attempt, the researcher will enter the assessment area and stop the recording devices. The student will return to choir class and another will enter the assessment area.

Two weeks following the conclusion of the five-week sessions, student participants will sight-sing short melody displayed on an iPad to determine what, if any gains were made in sight-singing achievement. All sight-singing attempts will be scored for accuracy by a panel of experts.

Following the conclusion of all data collection, Adam White will explain the research design in detail an allow for the opportunity to ask questions. You or student participants may still request removal from the study at this time and all data will be deleted.

Is there any way being in this study could be bad for my child?

There is nothing bad that will happen to your child though they may feel nervous about engaging in sight-singing activities. You or your child may decide to drop out of the study at any time and by notifying your teacher, Sarah Bartolome, or Adam White.

If you say that you do not want your child to be in this research?

Participation in research is voluntary. You can decide you do not want your child to participate in this research and it will not be held against you or your child in any way.

You can say "Yes" but change your mind later:

You can have your child stop and leave the research at any time and it will not be held against your or your child. We can end sight-singing sessions at any time and your child can select alternative sight-singing procedures provided by your teacher. Just let me know if you want to do this. If this happens, I will ask if any sight-singing recordings collected up until that point may be used in the research.

What happens to the information collected for the research?

All sight-singing recordings will be shared with your teacher then coded for anonymity. Every attempt will be made to ensure no one besides the researcher, a panel of expert graders, and your teacher will ever see or hear the recordings. We cannot promise complete secrecy. Organizations that may inspect and your information include the IRB, other representatives of Northwestern University, or school/district personnel.

Here is some other information that is useful for you and your child to know:

Be aware that under the Protection of Pupils Right Act 20 U.S.C. (c)(1)(A), you have the right to review a copy of the materials that will be used with your students. If you would like to do so, please contact Sarah Bartolome, sarah.bartolome@northwestern.edu to obtain a copy of the materials.

Here is who you and your child can talk to:

If you have any questions, concerns, or complaints, you can talk to the Principal Investigator, Sarah Bartolome at 847-491-8948 or sarah.bartolome@northwestern.edu. You may also contact the student investigator, Adam White at 785-845-5520 or adamwhite2020@u.northwestern.edu. This research has been reviewed and approved by an Institutional Review Board ("IRB"). You may talk to them at 312-503-9338 or irb@northwestern.edu if:

- Your question, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research participant.
- You want to get information or provide input about this research.

Your signature documents your permission for the named child to take part in this research.

____ _______________
Signature of child Date

____
Printed name of child

____ _______________
Printed name of parent [ ] or individual legally authorized [ ] to consent for the child to participate Date

____ _______________
Signature of parent [ ] or individual legally authorized [ ] to consent for the child to participate Date

If signature of second parent not obtained, indicate why: (select one)

[ ] The IRB determined that the permission of one parent is sufficient.

[ ] Second parent is: [ ] deceased [ ] unknown [ ] incompetent [ ] not reasonably available

[ ] Only one parent has legal responsibility for the care and custody of the child

Appendix F, Sight-singing Melodies

Pre/posttest melodies

Melody A

Melody B

Session melodies

Week 1

Week 2

Week 3

Week 4

Week 5

All melodies were selected from McGill, S., & Stevens Jr., H. M. (2003). *90 days to sight reading success: A singer's resource for competitive sightsinging.* Houston, TX: AMC Publications.