ABSTRACT

Title of Document: THE DEVELOPMENT AND VALIDATION OF AN INSTRUMENT TO MEASURE WIND ENSEMBLE ERROR DETECTION SKILLS AMONG INSTRUMENTAL MUSIC EDUCATORS

Karen Michelle Koner, Doctor of Philosophy, 2012

Directed By: Dr. Michael Hewitt, School of Music

The *Instrumental Music Error Detection Test*, or *IMEDT*, was developed to examine error detection ability regarding pitch rhythm and articulation errors in recordings of a wind ensemble. This test was designed to simulate an authentic rehearsal situation. The musical excerpts were selected from grade three band literature and performed with full instrumentation. A total of 30 errors was inserted into the recordings; 12 pitch errors, 12 rhythm errors, and 8 articulation errors. A university wind ensemble recorded the excerpts, first as written, or what was considered to be a “model

performance," and a second time with the errors inserted. The completed *IMEDT* contained two recordings of each of the eight musical excerpts, the first as written and the second with inserted errors.

The *IMEDT* was administered in six different test administration variations to determine the method that was most valid and reliable and had the highest internal consistency. Each test was administered in an individual setting with the participant and me and took approximately 45 minutes to an hour to complete. Sixty two participants completed this first phase of test administration. Using Cronbach's alpha to estimate the reliability and internal consistency, it was empirically decided that the test administration variation of score and recording with non-controlled time (S&R/N) had the highest alpha level. The order of musical excerpts was also determined empirically through this statistical test. Twenty additional participants completed the second phase of test administration of the *IMEDT* in the S&R/N method, again in a individual setting, taking approximately 40-45 minutes to complete. After data collection was complete, it was determined that the *IMEDT* was both reliable and internally consistent ($\alpha = .72$).

THE DEVELOPMENT AND VALIDATION OF AN INSTRUMENT TO MEASURE WIND ENSEMBLE ERROR DETECTION SKILLS AMONG INSTRUMENTAL MUSIC EDUCATORS

By

Karen Michelle Koner

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Dr. Michael Hewitt, Chair
Dr. Bruce Carter
Dr. Kenneth Elpus
Dr. Sherrick Hughes
Dr. Michael Votta
Dr. Trevor Parry-Giles

Acknowledgements

The process of completing doctoral work and a dissertation is a difficult journey. I could never have made it through without the love and support of my friends and family. A special thanks to my little sister Ellen Koner for all the support and encouragement when I felt like giving up over the past year. I would also like to thank Autumn Palmer and Jaymonde Errico for the support, phone calls, laughter, and being the two best friends a girl could have.

Thank you to Elizabeth Yackley and Carolyn Sweterlitsch for all the assistance with data collection and editing. I never would have made it through this dissertation process without the two of you. Additionally, I would like to thank the six gentlemen on my dissertation committee for being supportive, positive, and actually a joy to work with. Finally, a very special thank you to Michael Hewitt for four years of mentoring, advising, listening, editing, guidance, friendship, support, and fabulous summer camps!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Music educators must have aural diagnostic skills to run an effective rehearsal. It is imperative that the educator continuously recognizes different types of errors while students perform in order to assist with student improvement during rehearsals and performances (DeCarbo, 1982; Sheldon, 1998). An efficient and effective rehearsal, as well as a musically expressive performance, may be dependent upon the music teacher's error detection ability, in both speed and accuracy (Brand and Burnsed, 1981; Byo, 1993; Crowe, 1996). This ability to assess the ensemble, detect errors, and correct errors is a fundamental component of teaching and learning in music education. Proficiency in error detection is imperative to be able to assess a musical performance (Sheldon, 2004) and can be linked to exceptional teaching (Doerksen, 1999).

The importance of error detection ability has been investigated and discussed among scholars (Byo, 1997; DeCarbo, 1982; Doane, 1989; Doerksen, 1999; Forsythe & Woods, 1983; Grunow, 1980; Hochkeppel, 1993; Ramsey, 1979; Sheldon, 1998; Stuart, 1979; Taebel, 1980; Waggoner, 2011). Taebel compiled a list of competencies that in-service music educators felt were important concerning pupil learning in instrumental, choral, and general music. The primary musical competency rated among, instrumental, choral, and general music educators was aural skills ability. Taebel defined aural skills as the ability to "detect errors or problems in musical performance" (p. 189).

Influences on Error Detection Ability

Music education students receive training in aural skills during their undergraduate classes, during music theory or aural skills classes where the focus is often on learning and practicing sight singing along with melodic and harmonic dictation, by exhibiting the ability to take aural dictation (NASM, 2012). Aural skills training often have no component devoted to error detection, and there is no instructional method for teaching students to detect pitch and rhythm errors in an ensemble (Larson, 1977; Sidnell, 1971). Therefore, with no specific error detection training in the current music teacher educator curriculum, where do pre-service and in-service music educators receive this training or, if it is not a skill that can be taught in the classroom, what other experiences may impact this ability?

Brand and Burnsed (1980) examined the specific music abilities and experiences of pre-service music educators to try and predict error detection ability. The participants, 21 undergraduate instrumental music education majors, completed the *Music Error Detection Inventory* test designed by the researchers. Brand and Burnsed examined correlations between the participants' error detection test scores and the number of instruments they played, ensemble experience, ability in music theory, sight singing and ear training, and years of private instrumental instruction prior to entering college. They discovered that there was no statistically significant correlation between error detection ability and any of the five variables.

Scholars have also examined whether different musical settings can impact a pre-service or in-service music educator's ability to determine musical errors. Sheldon (2004) determined that 90 undergraduate music education students identified errors with higher accuracy when the errors were made in the soprano voice of a multi-voice texture and also most accurately during the first listening of a piece. The students correctly identified the least amount of errors, when the errors occurred in the lower voices and during the third listening of a piece. Byo (1997) examined the impact of texture on error detection ability and found that undergraduate and graduate students in music education scored significantly higher in musical excerpts with one-part as opposed to excerpts with two- or three-parts. He also found that graduate and undergraduate students in his study had a cumulative correct response rate of less than 50% in regards to error detection scores indicating that error detection is perhaps a weak skill for musicians and music educators. In addition, as the number of parts in the music increases, the error detection scores tend to decrease (Byo, 1993, 1997; Sheldon, 1998), indicating that music educators and conductors may be less likely to find errors in a large ensemble setting than in a single section or instrument setting.

Instrumental Error Detection Tests

Before exploring the *Instrumental Music Error Detection Test*, *IMEDT,* and what influenced and impacted its development, it is important to examine other error detection tests that have been used in music education research and are still currently in use. Byo (1993) developed a stimulus audiotape for error detection. Using grade four (of six) band literature, he reduced two full band scores to four voices and also transposed all instrumental parts to the concert key, with one excerpt using homophonic texture and the others polyphonic. Pitch and rhythm performance errors were inserted into the band scores in order to create an audio recording of the music being performed with intentional errors. To develop the audiotape, a digital synthesizer was used to record the excerpts under two conditions, single timbre (piano) and multi-timbre (clarinet, soft brass, flugelhorn, and contrabass). Participants studied each score for one minute, listened to each error recording three times, and circled errors in the score while marking an “R” for rhythm error and a “P” for pitch error. Each recording contained anywhere from zero to three errors.

A similar process was followed in developing a stimulus tape for a later study that examined the effects of texture and number of parts on the ability to detect errors (Byo, 1997). To develop the stimulus audiotape for error detection, Byo selected band literature that was polyrhythmic and homorhythmic in a one, two, or three part texture. Using the music notation software *Finale* (version 2.6), a digital synthesizer, and a stereo cassette deck, twelve errors were inserted into four musical excerpts. Participants were given scores in concert keys with no tempo, dynamic, or interpretation markers to interfere with detecting rhythm and pitch errors that were inserted into the recordings. These recordings were created with a piano timbre using a digital synthesizer. Again, participants circled errors in the score while marking an “R” for rhythm error and a “P” for pitch error.

Sheldon (1998) used test development methods derived by Byo (1997), to determine whether error detection ability was impacted by contextual sight singing and aural skills training. She used a similar testing method in a later study (2004); however in this setting the participants were allowed to listen to a correct aural example before the altered examples (2004). These examples consisted of four parts, utilized synthesized sounds from a MIDI keyboard and contained a combination of woodwind and brass timbres including oboe, flute, English horn, clarinet, bass clarinet, bassoon, trumpet, horn, trombone, euphonium, and tuba. The examples contained 120 errors, 20 of each type (articulation, dynamics/balance, intonation, pitch, rhythm, and tempo) dispersed through 12 musical examples and divided between all four voice parts.

The PRAXIS exam series, the primary exam series for teacher certification in 47 states, contains a component devoted to error detection ability. Pre-service educators typically complete several PRAXIS tests prior to applying for a teaching license and teacher certification. One of the PRAXIS test measures basic skills in reading, writing and mathematics while the second test examines specific musical content and pedagogical knowledge (http://www.ets.org/praxis). The content knowledge test for potential music educators includes an error detection component. This component requires that test takers listen to a recording of an instrumental and a choral ensemble and identify errors in regards to balance, accents, articulation, dynamics, and tempos. Test takers are not to indicate errors concerning pitch and rhythm and are penalized if they do (Educational Testing Services, 2003).

The *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979) has been used in music education research to examine error detection ability (Doane, 1988; Forsythe & Woods, 1983; Hochkeppel, 1993; Van Oyen and Nierman, 1998). The MLR was derived from the *Visual-Aural Discrimination Skills* test, or VADS, developed by Grunow (1980). Even though the VADS was used in Grunow's research prior to the development of the MLR, the MLR was published one year prior to Grunow's research. To develop the VADS, Grunow compiled 63 compositions from the Julliard Repertory Library to represent a broad range of musical styles and periods encompassing over 400 years of Western music. The musical excerpts included two-, three-, four-, and five-line scores using string, woodwind, and brass ensembles. Each excerpt was four to sixteen measures with difficulty ranging from elementary to high school performance levels. Each excerpt was recorded by an ensemble at the University of Michigan as "acceptable" or with “discrepancies” (p. 18). Participants listened to each excerpt under the two conditions of “acceptable” and with “discrepancies,” and marked errors in a general music error category or a specific technical error category. Errors listed as “general music criteria,” included tempo, balance, style of articulation, tone quality, and intonation, and errors described as “specific technical criteria,” included rhythm, note and pitch accuracy, phrasing, articulation, dynamic contrast, and ensemble.

The reliability of the *Visual-Aural Discrimination Skills* assessment tool is unclear after two administrations that were used to examine the reliability of the error detection instrument (Grunow, 1980). Using split-halves reliability estimation, the first test administration resulted in an overall reliability coefficient of .66, with the general music criteria resulting in an alpha level of .40 and the specific technical criteria at .74, which did not meet the .70 threshold typically used as a reference point when examining the reliability of an instrument in the social sciences (Field, 2009). The second time the VADS test was administered; the reliability coefficient was .71, with the alpha level for general music criteria being .34 and the alpha level for the specific technical criteria being .81. Grunow (1980) also estimated stability of the VADS test through the test-retest method, however, no mention of establishing validity is discussed in the research.

The VADS grew into the *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979) keeping the same error types and categories described above. In addition, the MLR contained excerpts with two-, three-, four-, and five-parts, fewer than the number of parts often used in a full wind ensemble. Also, as the number of musical parts increase, the number of musical examples decrease, with approximately five times more examples in two parts than in five parts. This could impact data and scores obtained from this test since it has been determined that as the number of parts increase the ability to detect errors decreases (Byo, 1993, 1997; Sheldon, 1998).

Rationale

While the tests that have been designed to examine error detection ability hold value in research and music teacher training, I sought to create an error detection test that would not only help music educators determine errors in a setting that was as authentic to an instrumental music rehearsal as possible, but also to serve as a valid and reliable tool in determining a music educator's error detection ability. I developed a test that used recordings of a full wind ensemble, rather than using music created electronically or via compositional software, as in previous studies (e.g., Byo, 1993, 1997; Grunow, 1980; Sheldon, 1998; Sidnell, 1971). The errors inserted into the recordings consisted of measurable errors of pitch, rhythm, and articulation. While I acknowledge that there are possible discrepancies in other areas of the ensemble, such as balance, blend, and intonation which can be distracting to the listener, I did not include these as part of the test for I desired to develop an error detection test that was practical, reliable, and valid in determining an instrumental music educator's error detection ability.

Pitch and rhythm errors were included on the *IMEDT*, as they are described as "typical errors" encountered during a rehearsal (Byo, 1993; Ramsey, 1979). In addition, Sheldon (2004) discovered that pitch, rhythm, and articulation errors are more correctly identified over tempo, dynamics/balance, and intonation errors. Furthermore , Grunow (1980) states that the low reliability coefficients seen in the general music criteria section of the VADS test reflect the nature of the musical criteria being too subjective to indicate whether a performance was acceptable, unacceptable, or questionable. I did not want this subjectivity to impact the *IMEDT* therefore this musical criteria was not used to define errors in the recorded performances, and only errors in pitch, rhythm, and articulation were included.

Purpose Statement

The purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances.

Research Questions

This research sought to investigate the following questions:

1. What is the validity of the *Instrumental Music Error Detection Test* (*IMEDT*)?
2. What is the reliability of the *Instrumental Music Error Detection Test* (*IMEDT*)?
3. What is the optimal method for administration of the *Instrumental Music Error Detection Test* (*IMEDT*)?

Theoretical Framework

Classroom teaching is typically organized by goals and objectives that influence instruction and are assessed by a large event or exercise, which can be referred to as a classroom assessment event (Brookhart, 2007). In this theoretical framework, the term classroom assessment event describes "the set of learning and assessment tasks assigned, standards and criteria set, and feedback given that constitute the formal presentation" (p. 166). These events may vary in size and may be repeated events, in which the experiences become a pattern that is known as the classroom assessment environment.

Error detection can be used as an imperative teaching tool in the classroom assessment environment of an instrumental music classroom. A music educator must continuously recognize different types of errors while students perform in order to assess student improvement during rehearsals and performances (DeCarbo, 1982). The idea of perceiving, diagnosing, and correcting musical behaviors (Brand & Burnsed, 1981) can be seen as a repeated event that shapes the classroom assessment environment. Music educators assess students and provide feedback to prepare students for a school concert, music festival, or music contest. This classroom assessment environment is a cycle of rehearsing, assessing, rehearsing, assessing, etc., which eventually leads to the large event. This theoretical framework regarding the use of error detection in a classroom assessment environment can be seen in Figure 1.

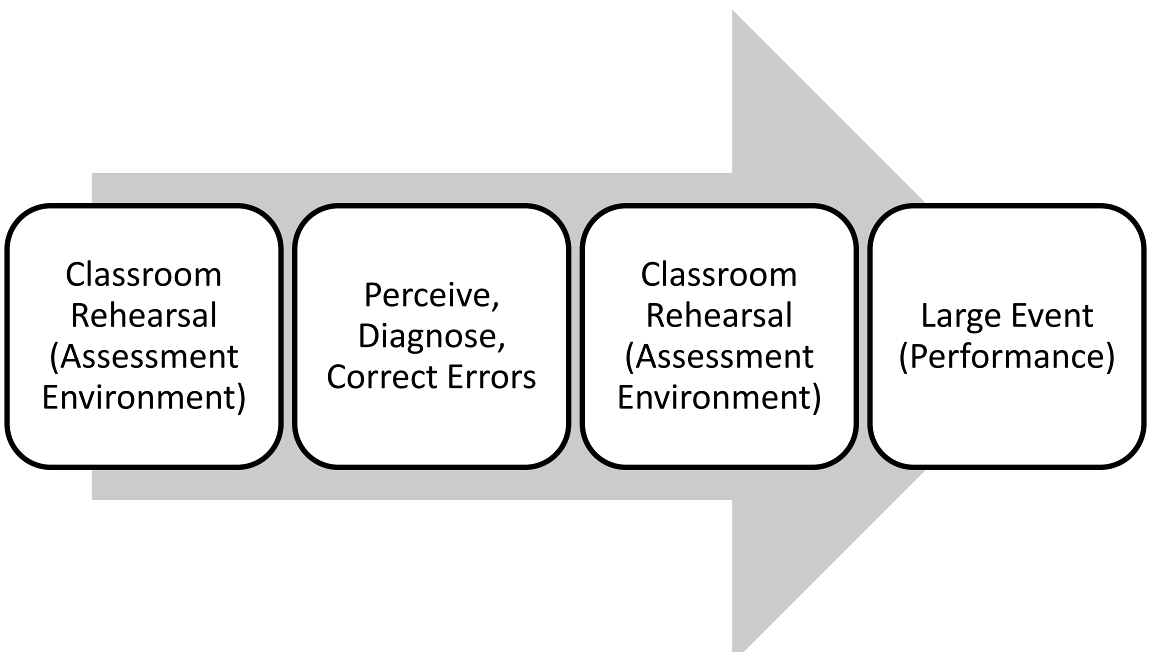

Figure 1: Theoretical Framework of Error Detection in Instrument Music Classrooms as an Assessment Tool

Definitions

Error detection

For the purpose of this study, error detection was defined as demonstrating aural and visual discrimination ability. This ability was assessed by identifying aural performance errors while examining the accompanying written musical score (Hochkeppel, 1993) in one of three conditions (a) studying the written musical score prior to listening to the performance with inserted errors; (b) not studying the written musical score but listening to a model performance of the musical excerpt prior to listening to the performance with inserted errors or; (c) studying the written musical score and listening to a model performance of the musical excerpt prior to listening to the performance with inserted errors. These three conditions were administered to participants with either controlled time or non-controlled time.

Error score

Each measure of music was labeled an item. The combined musical excerpts contained 82 measures, of which each was considered an item. An item was considered to be correct if it was identified as an error in a measure that had an error or identified as no error in a measure that had no error. Correct answers received one point. The *IMEDT* contained one measure in which two errors occurred. If only one error was identified a score of .5 was given and if both errors were identified the full point was given. In addition, a participant would receive no point if they identified an error in a measure where there was none (a misidentified error). Additionally, if a participant identified an error correctly, but also listed a misidentified error in the same measure, a score of .5 was given.

Nature descriptive score

The participant's nature descriptive score was determined by an item analysis on a scale of 0 to 30. In this item analysis, the participant would receive a half point (.5) if he or she correctly identified the instrument or instruments that were making the error (e.g., alto saxophone two and clarinet two) and another half point (.5) if the actual error was determined (e.g., alto saxophone two and clarinet two are playing a concert Eb instead of concert E natural). Therefore on this item analysis, a participant could receive a zero, .5, or a one for each item. If the nature of all items was determined, the highest score for the nature descriptive score was 30.

Pitch errors

A pitch error was defined as a pitch being performed at a minor or major second above or below the written pitch (Byo, 1997; Sheldon, 1998). This score was calculated from the error score. A participant could receive a zero if he or she did not indicate a pitch error when one occurred or a one if an error was indicated when it occurred. One point was given if the participant accurately indicated the measure, beat, and type of error that occurred. This information was determined only by the error score and was not impacted by the nature descriptive score. Pitch error scores could range from zero to 12, since there were 12 pitch errors in the *IMEDT*. These errors are described in Appendix C.

Rhythm errors

Rhythm errors were created by manipulating the original rhythms in one of two ways, either through elongation or diminution of the original note value or by performing the rhythm patterns earlier or later than written. This score was determined from the error score above. A participant could receive a zero if he or she did not indicate a rhythm error when one occurred or a one if he or she did indicate the rhythm error when it occurred. This score was calculated by participants accurately indicating the measure, beat, and type of error that occurred. This information was determined only by the error score and was not impacted by the nature descriptive score. Rhythm error scores could range from zero to 12, since there were 12 rhythm errors in the *IMEDT*. These errors can be found in Appendix D.

Articulation errors

Articulation errors were defined by manipulating slurred and/or articulated passages. Passages that were originally written as slurred were articulated and passages that were written as articulated were slurred. This score was determined from the error score above. A participant could receive a zero if he or she did not indicate an articulation error when one occurred or a one if he or she did indicate the articulation error when it occurred. This score was calculated by participants accurately indicating the measure, beat, and type of error that occurred. This information was determined only by the error score and was not impacted by the nature descriptive score. Articulation error scores could range from zero to six, since there were six articulation errors in the *IMEDT*. These errors can be found in Appendix E.

Student teaching interns

For the purpose of this study, student teaching interns were included in the study's sample. These interns were undergraduate instrumental music education students who had completed all coursework and were completing a final internship. The College of Education at the University of Maryland defines this internship as a full-time commitment. They state that "all core requirements and pre-professional requirements must be completed prior to the internship (University of Maryland College of Education website: http://www.education.umd.edu/EDCI/info/internship.html).

Wind ensemble

In the early 1950s, Frederick Fennell developed a new ensemble, which he wanted to be "a carefully-balanced instrumentation capable of performing styles from the 16th century and moderate-sized chamber music to Paul Hindemith's new *Symphony in B-flat"* (Battisti, 2002, p. 54). This ensemble type described by Fennell is similar to the ensemble that created the recordings for this study. These recordings were made by a university wind orchestra with full band instrumentation which contained approximately 50 undergraduate and graduate music majors. The full list of instrumentation used in these recordings can be found in Appendix A.

Overview of Remaining Chapters

This chapter served as an introduction to the current study, which sought to develop an instrument to measure errors detection skills among music educators and determine the most valid and reliable method in which to administer the instrument. Chapter two is a review of the related literature, which will be discussed in three sections; (a) contributing variables to error detection; (b) music assessment tests and; (c) validity and reliability of error detection programs and testing materials. Chapter three describes the methodology and design of the instrument used in the study while results are presented in chapter four. Finally, chapter five discusses any conclusions and implications this study may have on current music educator practices and training.

Chapter 2: Review of Literature

The purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances. The review of literature pertaining to the topic of error detection was divided into three sections, contributing variables to error detection ability, music assessment tests, and validity and reliability of error detection programs and testing materials. The first section, contributing variables to error detection ability was divided into three subsections; (a) musical setting of errors; (b) aural skills and; (c) score study. The second section, music assessment tests, was divided into three subsections; (a) musical aptitude tests; (b) musical achievement tests and; (c) music performance assessments. The third body of literature, validity and reliability of error detection programs and testing materials was divided into five subsections; (a) *Score Reading Ability Test* (Sidnell, 1971); (b) *Test in Error Detection Ability* (Ramsey 1978, 1979); (c) *Conducting and Written Test* (DeCarbo, 1981, 1982); (d) *Pitch Error Detection Test* (Malone, 1985) and; (e) The *Visual – aural Discrimination Skills* and *MLR*: *Instrumental Score Reading Program* (Froseth & Grunow, 1979; Grunow, 1980).

Contributing Variables to Error Detection Ability

Musical setting of errors

The ability to detect a large variety of errors in a performing ensemble is an essential skill of a music educator (Brand and Burnsed, 1981; Doerksen, 1999; Taebel, 1980). One method for classifying the types of errors heard in a rehearsal are general music errors and specific technical errors, which were categories used in the development of the *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979). General music errors are those that involve the tempo of the piece, balance of the ensemble, style of articulation, tone quality, and intonation, while specific technical errors include correct rhythms, note and pitch accuracy, phrasing, articulations and ornamentations, and dynamic contrast. While this present study is examining only specific technical errors, the similarities and differences in examining errors from both the general music category and the specific technical category will be discussed.

Blocher (1986) attempted to determine whether a hierarchy exists in which college band instrumentalists determine common performance errors, including articulation, dynamics, intonation, note accuracy, phrasing, and rhythm accuracy. By randomly assigning 141 instrumentalists enrolled in wind orchestra, symphonic band, concert band, and jazz ensembles to conducting or non-conducting conditions and implementing an error detection test, which was created with randomly assigned errors in the areas of articulation, dynamics, intonation, note accuracy, phrasing, and rhythm accuracy to eleven recorded brass trios, he determined that not only did increased conducting have no impact on error detection skills, but also there was no significant difference in overall error detection performance in relationship to the type of error selected. From this data, he concluded that there was not a hierarchy in determining types of errors in an ensemble performance.

In contrast, other researchers have discovered that there is a slight hierarchy, with participants finding rhythm errors with greater accuracy than pitch errors (Byo, 1993;

Hopkins, 1991; Sheldon, 1998). In a study of 60 graduate and undergraduate music education majors, Byo (1993) developed a stimulus audiotape for error detection. Using grade four band literature, he reduced two full band scores to four voices in a concert key, with one excerpt being homophonic and the others polyphonic. To develop the audiotape, a digital synthesizer was used to record the excerpts under two conditions, single timbre (piano) and multi-timbre (clarinet, soft brass, flugelhorn, and contrabass). These recordings were adapted to include 32 intentional errors, 16 rhythm and 16 pitch errors. After allowing graduate (n = 20) and undergraduate (n = 40) students to listen to each recording three times, he determined that not only were participants more accurate in determining rhythm errors, but were most accurate in finding these rhythm errors in the bass and tenor voices in music with a single timbre setting. In a multi-timbre setting, subjects were more accurate in finding errors in the soprano voice. Timbre had no effect in rhythmic error detection ability in regards to the alto voice.

Locy (1996) implemented *Gordon's Instrumental Timbre Preference Test* (1984) to help determine if undergraduate students' timbre preference and instrumental timbre influenced their ability to detect errors. Using Gordon's test as well as a test developed by the researcher, Locy recruited 147 conducting students in 11 colleges and universities to participate in these tests and this study. He discovered that not only does the timbre preference have no impact on the students' ability to detect errors, but the timbre of the melody performing the error also did not impact the error detection ability of undergraduate students.

Sheldon (2004), using 90 undergraduate music education majors from two large NASM-accredited universities, gave students three opportunities to listen to musical excerpts with inserted errors; however, her participants were able to listen to a correct aural example before the altered examples. These examples were four part, synthesized sounds from a MIDI keyboard and contained a combination of woodwind and brass timbres including oboe, flute, English horn, clarinet, bass clarinet, bassoon, trumpet, horn, trombone, euphonium, and tuba. These examples contained 120 errors, 20 of each type (articulation, dynamics/balance, intonation, pitch, rhythm, and tempo) dispersed through 12 musical examples and divided between all four voice parts. Sheldon determined that students labeled errors in voice one, the soprano voice, more accurately than errors in the lower voices. In addition, participants labeled the greatest number of errors during the first listening and the fewest during the third listening. Sheldon examined errors in articulation, dynamics/balance, intonation, pitch, rhythm, and tempo. She determined that participants identified errors in pitch, articulation, and rhythm with much greater accuracy than errors in tempo, dynamics/balance, and intonation. In relationship to the categories determined by Froseth and Grunow (1979), students were more successful in determining specific technical errors than general music errors.

Musical texture has also been found to contribute to a music educator’s ability to determine pitch and rhythm errors in musical examples (Byo, 1997; Crowe, 1996; Sheldon, 1998). Subjects scored significantly higher in identifying pitch errors in one-part musical excerpts, than on two- or three-part examples (Byo ,1997; Sheldon 1998). One hundred and fifty graduate and undergraduate music majors in Byo’s (1997) study received similar scores in detecting errors for musical excerpts with one- and two-part settings while listening to a stimulus audiotape; however, there was a significant drop when the music was in a three-part setting. In addition, these homorhythmic examples, defined as “music having a melody in the top voice that is supported by vertically oriented accompanying parts” (p. 55), gained slightly higher error detection scores by subjects than music that was polyrhythmic, defined as excerpts that featured a rhythmic variety in simultaneous parts. The highest error detection scores in this study were seen in two-part homorhythmic excerpts and lowest on two-part polyrhythmic excerpts. Byo’s sample, the graduate students (*n* = 45) scored with significantly higher accuracy on error detection than the undergraduate students (*n* = 105); however, overall in the study, music students, graduate and undergraduate, had a cumulative correct response rate of less than 50%.

Waggoner (2011) also found conflicting results examining 18 undergraduate music education majors. Participants listened to performances of band literature containing pitch and rhythm errors under two conditions, on recordings and while conducting a live ensemble. Half of the performances were with the full ensemble, while the other half were with a single section of the ensemble. Participants completed the task of detecting errors in the ensemble under both conditions, therefore serving as their own control group. Similar to previous studies (Byo, 1993; Hopkins, 1991; Sheldon, 1998), Waggoner found that rhythm errors were identified much more successfully in musical selections of single textures, as opposed to musical selections with a full ensemble texture. However, Waggoner also discovered that pitch errors were identified more correctly in a full ensemble texture, as opposed to a single texture setting.

Summary

The ability to detect errors in a performing ensemble is an essential skill of a music educator (Brand and Burnsed, 1981; Doerksen, 1999; Taebel, 1980). It has been discussed that musicians determine errors such as correct rhythms, note and pitch accuracy, phrasing, articulations and ornamentations, and contrast in dynamics with greater accuracy than errors such as tempo of the piece, balance of the ensemble, style of articulation, tone quality, and intonation in a musical ensemble (Byo, 1993, 1997; Sheldon, 1998, 2004; Waggoner, 2011). In addition, researchers have discovered that participants found rhythm errors with greater accuracy than pitch errors (Byo, 1993; Sheldon, 1998). In regards to the musical setting, it appears that when the number of parts in the music increases, the error detection scores decrease (Byo. 1993, 1997; Sheldon, 1998). This suggests that music educators and conductors are less likely to find errors in a large ensemble setting then in a single section or instrument setting.

Aural skills

Larson (1977) states that aural skills instruction for music majors at the college level "often attempts to develop aural discrimination abilities through experiences in dictation and sight singing, to the neglect of abilities to detect performance errors" (pp. 264). However, the ability to detect errors and aural skills development are paramount to a music educator (Brand & Burnsed, 1981; Doerksen, 1999; Taebel, 1980). As stated in chapter one, the primary musical competency rated by public school music educators that is needed to be a successful music educator was aural skills (Taebel). Throughout this next body of literature, the development of error detection skills in relationship to aural skills development and abilities will be discussed and explored.

Harrison, Asmus, and Serpe (1994) conducted a study regarding freshman undergraduate music theory students, including aural skills development. They defined aural skills in two parts, ear training and sight singing. Ear training was defined as the process students use to identify chords, intervals, rhythms, and other musical ideas. Sight singing was defined as the skills to sing a melodic line upon first exposure to the melody. For the purpose of this study and this body of literature, aural skills will be defined by the explanations provided by Harrison, Asmus, and Serpe.

Larson (1977) sought to examine the competency of undergraduate music majors in three areas of aural skills, melodic error detection, melodic dictation, and melodic sight singing. Using 204 undergraduate music majors, Larson picked 12 melodies from published sight singing text books that were diatonic, chromatic, and atonal. After administering group tests in error detection and melodic dictation and individual tests in sight singing, he discovered there was a higher relationship between melodic dictation scores and error detection scores than between sight singing scores and error detection scores, regardless of what type of melody was being examined.

Killian (1991) examined the relationship between sight singing and error detection in junior high school students and discovered similar results. In comparing 75 junior high choir students' sight singing accuracy and ability to perceive errors in recorded examples, Killian assigned three choir classes a sight singing task as a regular part of the class requirements. These students were familiar with both notation and solfege syllables. In addition, a perception test was designed with a voice and a piano performing unison pitches. However, throughout the test, the vocalist would perform a different pitch then the piano. Participants were asked to sight sing (perform) and complete the Error Test (perception). Killian discovered that for students who scored high or medium on sight singing (performance), there was no significant difference in the correlation between their sight singing ability (high or medium) and their error detection ability. However, low scoring singers were significantly more accurate on error detection tasks.

In contrast, Sheldon (1998) found that additional sight singing training did increase students' error detection abilities. Thirty undergraduate music education majors received identical training in instrumental methods and conducting, however half of the students (n = 15) received an additional contextual sight-singing and aural skills training of 50 minutes per week over an 11 week period. Undergraduate students with the additional training were more accurate in error detection and less likely to assume errors when there weren't any errors happening in the musical selection. These students performed better at detecting rhythm errors compared to pitch errors and performed best in detecting errors in a one-part musical setting, as opposed to a two- or three-part setting.

Summary

It appears that increased aural skills training can improve error detection ability (Larson, 1977; Sheldon, 1998). Larson indicates that the skills needed to be successful in error detection are more closely related to a student's ability to perform well in melodic dictation than in sight singing. The relationship between sight singing ability and error detection ability in middle school students also indicates that perhaps there is not a strong correlation between sight singing and error detection (Killian, 1991). Sheldon determined that increased contextual aural skills training did increase error detection scores; however, she states that this additional training may only increase short term error detection ability. Based on this body of literature, it is still unclear the extent of the relationship or correlation between aural skills and error detection ability. The next section of literature, score study, will address the impact of different methods of score study on error detection ability.

Score study

In the book *Guide to Score Study for the Wind Band Conductor*, Battisti and Garofalo (1990) suggest that to become familiar with a score, a conductor or music educator must follow a four step process. The first step is score orientation, which includes examining specific information pertaining to the piece and a "cursory glance" at each page of the score (p. 4). The second step is score reading, which involves "the conductor's musical imagination, intuition, inner hearing ability, memory, and emotions" (p. 22). They believe there are two objectives to this second step, for the conductor to acquire an image of the music and to develop a musical feeling of how to express this image. The third step, according to Battisti and Garofalo, is score analysis, which is the systematic and thorough study of the musical score. Finally, the fourth step to score study is score interpretation, in which the conductor uses all of his or her knowledge from steps one, two, and three to create an interpretation of music. These steps are helpful when learning how to examine a musical score, however, how does each conductor carry out the four steps and does the method in which a conductor studies a score impact their ability to hear and correct errors when leading an instrumental ensemble rehearsal? The relationships between score study, conducting, and error detection skills will be examined and evaluated throughout this third body of literature.

Forsythe and Woods (1983) examined how conducting a score while also listening to the music being played would enhance error detection skills. They asked 20 undergraduate and 20 graduate instrumental music education majors to complete an error detection test which was adapted from the *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979). This test consisted of twelve recorded instrumental musical excerpts with various errors inserted. These errors were general music errors and specific technical errors (Froseth and Grunow, 1979). While listening to the musical examples on the test, participants conducted half of the examples while listening and listened to the other half of the examples without conducting. The researchers determined that there was no significant difference between undergraduate and graduate students in error detection scores. However, there was a significant improvement on error detection when students were not conducting the score as opposed to when students were listening while conducting the score.

There are different ways to examine and study a musical score and researchers have performed experiments in comparing different score study methods to determine which are most beneficial to developing or increasing aural diagnostic skills (Crowe, 1996; Grunow, 1980; Hochkeppel, 1993; Hopkins, 1991). Grunow compiled 63 compositions from the Julliard Repertory Library to represent a broad range of musical styles and periods encompassing over 400 years of Western music. The musical excerpts included two-, three-, four-, and five-line scores using string, woodwind, and brass ensembles. Each excerpt was four to sixteen measures with difficulty ranging from elementary to high school level. Each excerpt was recorded by an ensemble at the University of Michigan as "acceptable" or with "discrepancies" (p. 18). Participants listened to each excerpt under the two conditions and marked errors in the general music error category or the specific technical error category. Participants completed one of four methods of studying the scores on the Error Test developed by Grunow; studying the score only, studying the score with recorded examples, studying the recorded examples only, and no preparation. He found that there was no significant difference of error detection scores in relationship to the method the participant used to study the score. Additionally, he found there was no predictive relationship between years of teaching experience, the grade level an educator taught, and the highest degree earned and error detection ability. However, other scholars have determined that different methodologies of score study can impact error detection scores.

Crowe (1996) also investigated four methods of instrumental score study; no score study, studying the score alone, studying the score with a correct aural example, and studying the score at an electronic keyboard. Thirty members of a beginning undergraduate conducting class at three different Midwestern universities completed all four methods of score study and completed four tests, with musical excerpts played through a MIDI keyboard, to determine which method was the most effective to enhance error detection scores. He determined that students who performed score study with correct recorded examples scored significantly higher on error detection tests than students who studied the score alone and silently. Similarly to research discussed above (Byo, 1997; Sheldon, 1998), he also discovered that as the number of parts in a musical example increased, the error detection became more difficult for participants.

Hochkeppel (1993) used error detection as a tool to measure the effectiveness of four methods of score study; studying at a piano or keyboard, using recorded models, singing the score, and studying the score silently. Using 47 music majors, including six graduate students, assigned to the four different treatment groups, Hochkeppel modified the *MLR Instrumental Score Reading Program* (Froseth and Grunow, 1979), as well as the *Test in Error Detection* (Ramsey, 1979), and the *Advanced Measures of Music Audiation* (Gordon, 1989) and implemented a pre-test, treatment, post-test design. Each subject participated in seven treatment session in accordance to one of the four methods of score study. He found students who studied the score silently had significantly higher scores in error detection than students who studied while singing or using a keyboard instrument.

Using choral music, Hopkins (1991) also implemented four methods of score study to determine its effectiveness on undergraduate music students' ability to detect errors in choral music. Using pianists and non-pianists, Hopkins separated students into four groups to study musical scores; using a piano, using a recording, sight singing, and silent inspection over a period of four days. Overall, he found that participants detected rhythm errors with more accuracy than pitch errors and pianists received higher scores than non-pianists, however, this difference was not significant. He determined that there was a significant difference in error detection scores of students who used recordings to study a score over students using a keyboard piano instrument.

However, Van Oyen and Nierman (1998) found results in disagreement to the research above. Forty seven undergraduate music majors who had successfully completed one and a half years of music theory/ear training courses were randomly assigned to two experimental groups using a pre- and post-test design. One group performed extended score study while the other group performed extended score study with recorded examples. The extended score study group studied an instrumental music score for two minutes before listening to an audio example with errors, while the extended study group with recorded examples received the same two minutes along with an additional correct aural version of the recorded material. In addition, both groups were asked to conduct along with the music during the post test. Using the *MLR Instrumental Score Reading Program* (Froseth and Grunow, 1979), it was determined that there was not a statistically significant difference in error detection scores between the general music error type scores and the specific technical error type scores between the two groups. Also, there was no statistically significant difference found between the two groups in overall error detection scores.

Another methodology used during score study is to have the conductor sing through parts while studying the score. Byo and Sheldon (2000) examined 41 upper level undergraduate music education majors' ability to detect errors while singing selected parts in a one-, two-, or three- part excerpt. Two stimulus tapes were developed including a total of 54 purposeful pitch and rhythm errors. Students were required to learn a musical score, which they defined as "being able to demonstrate the ability to sing all parts of each excerpt with a high degree of pitch accuracy" (p. 26). After students could sing the musical score at this level, they participated in an error detection test that either included them singing a certain line of the score while listening for errors or not singing at all while listening for errors. Using a pre- and post-test design, they determined that the singing had little effect on pitch and rhythm error detection in a one-part setting, however, in a two – and three-part setting, the singing actually had a negative effect on the error detection scores, therefore indicating that singing one part while listening for errors in other parts does not positively impact error detection ability. However, this ability to sing all parts of the score accurately or "learning the score" resulted in an increased accuracy to detect pitch errors, but did not greatly impact the ability to detect rhythm errors. Also, as seen in studies discussed above, the error detection scores decreased as the number of parts increased.

Summary

Scholars have examined the effectiveness of different methods of score study on error detection scores. These methods of score study include studying a score with a piano or keyboard (Crowe, 1996; Hochkeppel, 1993; Hopkins, 1991), studying a score silently (Crowe, 1996; Grunow, 1980; Forsythe & Woods, 1983; Hochkeppel, 1993; Hopkins, 1991; Van Oyen & Nierman, 1998), studying the score with a recording (Crowe, 1996; Forsythe & Woods, 1983; Grunow, 1980; Hochkeppel, 1993; Hopkins, 1991; Van Oyen & Nierman, 1998), studying a score while singing (Byo & Sheldon, 2000; Hopkins, 1991), studying with no score and only a recorded example (Grunow, 1980), and not studying the score (Crowe, 1996; Grunow, 1980). However, scholars have determined mixed results when attempting to determine the most effective method on error detection skills. Studying a score with correct aural examples may help increase the conductor's ability to detect errors while rehearsing his or her own ensemble (Crowe, 1996; Hopkins, 1991). Additionally, the act of score study and analysis, regardless of what method is used, can also help increase error detection scores (Hochkeppel, 1993; Van Oyen and Nierman, 1998).

Battisti and Garofalo (1990) suggest that a conductor should examine the score silently before listening to any aural examples of the music. This method is also commonly used in the teaching and the training of pre-service music educators. Colwell and Hewitt (2011) state "don't listen to a recording right away, but rather allow the musical mind to develop the structure, mood, phrasing, tempo, and desired artistry" (p.

342). The idea of studying a musical score with recorded examples may be beneficial to a conductor's or music educator's ability to determine errors in the ensemble (Crowe, 1996; Hochkeppel, 1993; Hopkins, 1991; Van Oyen and Nierman, 1998) and perhaps future investigation is needed to determine how beneficial this method of score study may be.

Musical Assessments

Assessment in music education began with the work of Seashore in 1919, examining musical aptitude. Error detection testing appears to have started with the work of Sidnell (1971) who implemented programmed instruction into a classroom setting to help improve and examine error detection ability. However, prior to this work, music education scholars had developed and used musical assessment testing to examine musical aptitude (Gordon, 1965, 1978, 1982, 1989; Seashore, 1919), musical achievement (Colwell, 1969; Gordon, 1970), and music performance (Watkins-Farnum, 1954, 1962; Farnum, 1969) of music students. These tests contain activities for students to complete (Colwell, 1969; Gordon, 1970; Seashore, 1919) that are similar to error detection tests used in the field of music education and therefore could be considered the predecessors that helped shape error detection testing.

Musical aptitude

The *Seashore Measures of Musical Talents* (Seashore, 1919) may be the most widely known published standardized musical aptitude test (Boyle & Radocy, 1987). Carl Emil Seashore believed that musical talent was an inherent gift of nature and could not be acquired. In addition, Seashore believed that musical talent was not just one particular attribute but rather consisted of a hierarchy of talents that worked together, meaning the term musical talent is used in a collective sense (Seashore, 1915). Humphreys (1993) states that Seashore was the most influential music psychologist of his era and inspired an intense interest in musical aptitude testing among music education researchers beginning in the 1920s.

Seashore's beliefs helped to shape his musical aptitude test, which was first published in 1919. This test initially contained six sections: pitch, loudness, rhythm, time, timbre, and tonal memory with a rhythm section was added to the test in 1925 (Boyle & Radocy, 1987). According to Boyle and Radocy (1987) a person who earned a high score on all sections under good listening conditions will likely be an excellent auditory discriminator.

Gordon developed The *Musical Aptitude Profile, MAP,* (1965) which was designed to assess musical achievement from the basic factors of musical aptitude; musical expression, aural perception and kinesthetic music feeling. This test contains three divisions; tonal imagery, rhythm imagery, and musical sensitivity. The test takes approximately 50 minutes to complete the test in its entirety and is intended for students in grades 4-12. Boyle and Radocy (1987) state that few published tests are more thorough than the *Musical Aptitude Profile*

Following the creation of the *MAP* (1965) Gordon created several additional musical aptitude tests; the *Primary Measure of Music Audiation, PMMA,* (1978), the *Intermediate Measures of Music Audiation, IMMA,* (1982), and the *Advanced Measures of Music Audiation, AMMA,* (1989). The three tests, all of which measure audiation were developed for different age groups of students; the *PMMA* for students in grades kindergarten through third grade (1965), the *IMMA* for students in grades first through fourth (Gordon, 1979), and *AMMA* for high school students (1989). Gordon (1989) defines audiation as being able to hear and comprehend music when the sound is not physically present. He stated that music aptitude is a product of both nature and nurture and believed that "one's musical aptitude will never reach a higher level than that with which one is born" (Gordon, 1989, p. 10).

Musical achievement tests

The *Musical Achievement Tests* (*MAT*) (Colwell, 1969) are designed to provide an accurate measure of musical achievement for important objectives in music education. The tests are designed to be useful for both the classroom teacher and music specialist. Colwell states that these tests are appropriate to use with any music textbook series and the information included in the *MAT* is vital to program planning, curriculum revision, and evaluation of objectives. These tests are designed to help improve teaching rather than assign grades to students on how well they score on the *MAT*.

According to Colwell (1969) the *MAT* has four purposes. The first is to determine how well each student has mastered basic auditory objectives from the school music program. It provides individual evaluation of each student in the classroom, as opposed to a group assessment. Teachers can use the *MAT* to obtain information on the specific strengths and weaknesses of each student in a music classroom. The second purpose of the *MAT* is to provide information concerning which students may profit the most from instrumental instruction. Colwell states that students who score high on the *MAT* have a good chance for success and mastery on an instrument. The third purpose is to inform teachers, administrators, parents, and students with information to help guide the student. Colwell states that the scores on this achievement test can help elementary schools plan curriculum for the musical level on the students. At the secondary level, the scores from the *MAT* may help guide students as to which music elective classes they would be best suited. A fourth purpose is to provide teachers and administrators with information to evaluate and improve the current music program. The test provides a comparison of means between individual classes, comparable schools, and past and current music programs.

The *MAT* (Colwell, 1969) contains two tests. Test 1, provides information regarding three musical skills; pitch discrimination, interval discrimination, and meter discrimination. These three sections do not require an ability to read music, but rather requires answers based on auditory responses. The pitch discrimination section measures the extent to which students are able to hear and identify musical pitches, intervals, and the direction a pitch moves (higher or lower). The interval discrimination section asks students to compare melodic notation with a presented recorded aural example and indicate the measures where the auditory and visual examples do not align. The third section, meter discrimination, also asks students to compare rhythmic notation with the aural example and indicate the measures that do not sound as written. Test 2, provides information on three musical skills; major-minor mode discrimination, feeling for tonal center, and auditory-visual discrimination for both pitch and rhythm. Students complete these sections identically to Test 1.

The *Iowa Tests of Musical Literacy* was developed by Gordon in 1970. The test contains six levels divided by difficulty. There are two primary sections, each divided into three subsections. The first primary section, Tonal Concepts contains a subtest, Aural Perception, which asks students to listen to melodies and classify the mode in which the music is performed (e. g., major or minor). The second subtest of Tonal Concepts is Reading Recognition, which requires students to detect discrepancies between an aural melody and written notation. The third subtest of Tonal Concepts is Notational Understanding which requires that students complete the notation for a melody that is presented aurally.

The second primary section of the *Iowa Tests of Musical Literacy* (Gordon, 1970), Rhythmic Concepts has the same subtests at the Tonal Concepts section. In Rhythmic Concepts, the Aural Perception subtest requires students to discriminate between meters. The subtest of Reading Recognition requires students to detect discrepancies when listening to aural examples and looking at the written rhythms. The final subtest, Notational Understanding, requires students to draw in notes, beams, rest, ties, and flags where appropriate.

Music performance tests

The Watkins-Farnum Performance Scale for Band Instruments-Form A and B (WFPS) (Watkins-Farnum, 1954, 1962) consists of fourteen musical excerpts for cornet, baritone in treble clef, clarinet (soprano, alto, and bass), saxophone, oboe, flute, horn, trombone, baritone in bass clef, bassoon, tuba, and snare drum. These fourteen musical excerpts are in a variety of key and time signatures. These excerpts get progressively more difficult from excerpt one to excerpt fourteen. The test must be administered individually. The instrumentalist performs for an adjudicator and is scored by the number of measures containing at least one error subtracted by from the total number of measures per excerpt. The adjudicator also indicates the type of error executed in each measure.

There are eight types of errors that can be heard and indicated on the *WFPS*. The first is a pitch error, which is defined as when the student adds an additional tone, omits a tone, or performs an incorrect tone. The second type of error is time, which is when a tone is not sustained for the correct length, either plus or minus one beat. The third type of error is a change of time error, which is heard when a tempo change is not performed or an incorrect tempo is performed. The fourth type of error is expression, which included when students do not observe dynamic markings and musical terms such as ritardando and crescendo. The fifth type is slur errors, which is when the student omits a slur, articulates a note that is indicated to be slurred, slurs a note which should be articulated, or breaks a slur in the middle. The sixth type of error is rests, when a student ignores a rest or fails to give it its correct value. The seventh type of error is holds and pauses, which can also be defined as rhythm errors. This type of error is defined as when a student pauses between measures or pauses between the notes in a measure. The eighth type of error is repeats, which is when a student fails to follow the repeat or fist and second ending sections in the music

The *Farnum String Scale* (*FSS*) (Farnum, 1969) was developed following *The Watkins-Farnum Performance Scale for Band Instrument* (Watkins & farnum, 1954, 1962). Farnum states that it was believed the *WFPS* would also be suitable for string students if transposed into concert keys and revised the test to create the *FSS.* This test is in the same format as the *WFPS* with similar types of errors to be detected by the adjudicator and a few additional error types added that are specific to string students. The error types of pitch, change of time, expression, rests, holds and pauses, and repeats are identical to the *WFPS*. However, Farnum added the types of errors of time and bowing.

He defined time errors as any note performed by the student that is not given the correct value. The bowing error section included a student using incorrect bowing, including bowing marking such as spiccato.

Summary

Musical assessment tests were the predecessors to error detection tests in music education and were designed to measure student's musical aptitude (Seahore, 1919; Gordon, 1965, 1978, 1982, 1989), musical achievement (Colwell, 1969; Gordon, 1970), and music performance (Watkins-Farnum, 1954, 1962; Farnum, 1969). These tests were designed to examine musical traits and qualities in music students. Error detection programs and testing materials, the third and final body of literature were typically designed to measure the skill of error detection in music educators rather than in the students they teach.

Validity and Reliability of Error Detection Programs and Testing Materials

When developing and administering programmed materials or testing in error detection it is imperative to use an instrument that is both valid and reliable. Creswell (2005) defines a reliable instrument as having scores that are stable and internally consistent. These scores should be similar when the instrument is administered multiple times, in different settings, or with different participants. In addition, the instrument is seen as internally consistent when a participant answers closely related questions from the instrument similarly.

Validity can be defined as using evaluative judgment to the degree in which empirical evidence supports the adequacy and appropriateness of generalizations based on test scores (Messick, 1989). This validity can typically be broken down into three categories: content validity, criterion-related validity, and construct validity. Content validity examines the extent to which the question on an instrument represents all questions a researcher may ask on a specific topic. Criterion-related validity examines if scores from an instrument can predict a particular outcome. Construct validity determines if scores from an instrument are meaningful, have a purpose, and can be used to understand the population being examined (Creswell, 2005). These definitions were used when examining the third body of literature, validity and reliability of error detection programs and testing materials.

Score Reading Ability Test

Sidnell (1971) discussed that there was no instructional method for the teaching of score-reading skill. He defined score reading as “the skill in the detection and identification of pitch and rhythm errors in instrumental performance” (p. 85). Students received aural skills in theory class but it was unclear how much these transferred to a conductor or teacher in front of an ensemble. He created a program that would develop self-instructional materials for students to use to develop error detection skills. For this program, he randomly selected musical examples that were from a standard repertoire list by the National Interscholastic Music Activities Council. Local middle and high school ensembles were then asked to sight read these selections while they were recorded and examined for common mistakes and errors in regards to rhythm, pitch, style, and intonation. However for the purpose of this research he chose to only include errors of the pitch and rhythm nature. Two hundred and forty errors were identified and the musical selections were recorded a second time by college performers, who intentionally inserted the errors. Instrumentation of only three, four, or five instruments was used and all recordings were compiled into 20 drill tapes, which contained 12 excerpts each. The musical scores used to accompany the tapes were all in concert keys to eliminate any transposing difficulties, in a four-line staff format, and in treble and bass clef only. Using a split-halves model, Sidnell found the reliability for this test to be a .93. The test was validated by members of the theory and conducting faculty at a university.

Twenty six undergraduate music education majors enrolled in a second semester of conducting were separated into two groups for ten weeks. During these ten weeks, the experimental group used the programmed materials Sidnell (1971) had developed, while the control group did not use any programmed materials. By having students implement these materials, he found that their error detection ability or what he referred to as “score reading skills” (p. 85), improved significantly. Following the positive results seen from the development of Sidnell’s programmed materials; other researchers began to develop different programs of their own.

Test in Error Detection Ability

Ramsey (1978, 1979) developed a test, *Test in Error Detection* (TIED), to examine if “significant gains in error detection ability, as measured by scores on the TIED, occur among college music students trained in pitch and rhythm error detection with programmed instruction using full-score band literature” (1979, p. 156). To develop the TIED, Ramsey selected full band literature of the medium difficulty level, grade three or four. From this literature he created a list of typical pitch and rhythm errors that can be seen when performing. A 30-piece wind ensemble recorded this literature and a panel reviewed the musical excerpts. An answer booklet was also created to accompany the test with three questions: 1. In what measure does the error occur?; 2. In what instrument does the error occur?; 3. What is the exact nature of the error?

A panel of three experts in instrumental music reviewed the musical excerpts chosen for the TIED. The panel was asked to listen to each excerpt that contained the error performances while following on a correct written score and asked to eliminate excerpts that included errors that were not seen as typical errors based on their own background and experiences. These errors, along with errors that were not perceptible were removed. Ramsey (1978, 1979) completed this process to ensure validity of the excerpts on the TIED.

Two estimates of reliability were used to examine the internal consistency and the stability of the TIED. In order to determine internal consistency, a split-half model was used on pretest data, which resulted in a reliability coefficient of .63. To examine the stability of the scores, Ramsey (1978, 1979) computed the Pearson product-moment correlation from the pretest and post-test scores, which resulted in a coefficient of reliability of .71.

A sample of 77 undergraduate music students was separated into four groups, receiving different amounts of instruction on the TIED, either six weeks, four weeks, two weeks, or no instruction. All students who participated in the programmed instruction made significant gains in error detection ability on the final post-test, while the control group who did not receive any programmed instruction did not make any gains in this area. Additionally, the group that received six weeks of training scored significantly higher than the groups who received two weeks of training or no training. However, the group that received four weeks of training only scored at a statistically significant difference in comparison with the control group who received no instruction.

Conducting and Written Test

DeCarbo (1981, 1982) did not find the same results with his programmed materials. Using a sample of 32 undergraduate music students, he created two groups, one receiving an additional 16 class sessions of conducting and one receiving 16 class sessions of programmed instruction. For the conducting group, eight compositions, from a variety of historical eras, including Renaissance, Baroque, Classical, Romantic, and Contemporary periods, with two-, three-, or four-parts, were performed and conducted in class. These same compositions and materials were used for the programmed instruction; however the students in this group did not get an opportunity to conduct the materials. After the 16 class sessions were complete, participants completed an error detection written and error detection conducting test. He found that there was no significant difference between written error detection scores between the two groups, however, the conducting group scored significantly better on the conducting Error Test.

All music used for the conducting class and for the programmed instruction was performed by an all brass ensemble and contained errors in dynamics, intonation, ensemble performance, note accuracy, rhythm accuracy, and style. During the recording session, the investigator, a professional conductor, the performers, and the recording engineer validated each of the musical examples. After the recording session, DeCarbo (1981, 1982) re-examined each performance to ensure all errors were present and no additional errors were made.

After the conducting and written tests were administered as a pilot study, DeCarbo (1981, 1982) examined the content validity and internal consistency between the written test and the conducting test by examining the data. Since the same tests were used in both the conducting and written situations, it was assumed that that they both had content validity. To estimate the reliability based on the internal consistency, a Kunder-Richardson Formula-20 was computed, which resulted in a reliability coefficient of .83 for the written test and .90 for the conducting test.

Pitch Error Detection Test

Malone (1985) also created a programmed instruction to increase pitch error detection ability, however these materials were developed for choral music education students, rather than the instrumental music education students. Pitch errors in this test consisted of intervals of perfect fifths, keys, chromatic pitches, intonation of sustained notes, repeated notes, and major thirds of a chord. Musical selections were chosen based upon these potential problems and performed by a small choral ensemble. Instructional materials developed for this test were implemented in three settings: one-on-one, small groups, and large groups. Four students completed the one-on-one teaching for these materials, ten students completed the small group instruction, and twenty-five students completed the large group instruction.

The Kunder-Richardson Formula-20 was used to compute the reliability coefficient of the pitch error detection test used for this program. The reliability coefficients for the intervals for perfect fourths and fifths was .95, for in the key was .95, for chromatic notes was .89, for sustained notes was .87, for repeated notes was .94, and for the major third of the chord was .54. To establish content validity on the *Pitch Error*

Detection Test, three experienced music educators examined the behavioral objectives of the test and decided each objective was reflected in the testing material. To establish concurrent validity, the participants' performance on the test was compared to performance on a professor designed test, where students viewed written choral music scores while listening to the excerpts being performed on the piano. Using Pearson's formula, a coefficient correlation of .92 was found. In addition, all recordings were examined by the investigator and another experienced choral conductor for performance quality in regards to style, tempo, rhythmic precision, balance, blend, intonation, tone quality, ensemble, and artistry and clarity of intended errors.

Overall, Malone (1985) found that students responded favorably to the programmed instruction. Students in the one-on-one group scored higher on pitch error detection, however it was unknown if this was a result of the programmed instruction or of the one person giving the instruction. The small group phase showed that several subjects scored below criterion-level on some items; however the large group scored significantly higher on the post-test indicating that the programmed materials were effective as measured by the Pitch Error Detection Test.

Visual – aural Discrimination Skills and MLR: Instrumental Score Reading Program

The *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979) was derived from the *Visual-aural Discrimination Skills* test, or VADS, developed by Grunow (1980). Grunow compiled 63 compositions from the Julliard Repertory Library to represent a broad range of musical styles and periods encompassing over 400 years of Western music. The musical excerpts included two, three, four, and five line scores using string, woodwind, and brass ensembles. Each excerpt was four to sixteen measures with difficulty ranging from elementary to high school level. Each excerpt was recorded by an ensemble at the University of Michigan under the two conditions of "acceptable" or with "discrepancies" (p. 18). Participants listened to each excerpt under the two conditions of acceptable and with discrepancies and marked errors in the general music error category or the specific technical error category. Errors described as "general music criteria," include tempo, balance, style of articulation, tone quality, and intonation, and errors described as "specific technical criteria," include rhythm, note and pitch accuracy, phrasing, articulation, dynamic contrast, and ensemble.

The VADS eventually grew into the *MLR: Instrumental Score Reading Program* (Froseth and Grunow, 1979) keeping the same error types and categories described above. This test has been used throughout research to investigate error detection ability amongst music educators, both pre-service and in-service (Doane, 1988; Forsythe & Woods, 1983; Hochkeppel, 1993; Van Oyen and Nierman, 1998). However, it is unclear of the reliability of this assessment tool. The VADS test had mixed results when examining the reliability of the test itself. The VADS test was administered twice (Grunow, 1980). Using split-halves reliability testing, the first test administration resulted in an overall reliability coefficient of .66, with the general music criteria resulting at an alpha level of .40 and the specific technical criteria at .74. This overall number is not quite above the .70 threshold typically used as a reference point when examining the reliability of an instrument in the social sciences (Field, 2009). The second time the VADS test was administered the reliability coefficient did surpass the mark of .70, by testing at a .71, with the alpha level for general music criteria being .34 and the alpha level for the specific technical criteria being .81.

Summary

It appears that programmed instruction in error detection creates positive results (Doane, 1988; Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971). In addition, the programmed materials and tests in error detection have proven to be overall reliable, internally consistent, and valid (DeCarbo, 1981, 1982; Malone, 1985; Ramsey, 1978, 1979; Sidenell, 1971). Researchers have most commonly used the split-half method (Grunow, 1980; Ramsey, 1978, 1979; Sidnell, 1971) and the Kunder-Richardson Formula-20 (DeCarbo, 1981, 1982; Malone, 1985) to determine if their test is reliable and internally consistent. Furthermore, when examining validity, it appears as if the most common method of supporting instrument validity is to have a panel of experts, such as university professors or other researchers, examine the instrument (Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971).

Implications

The ability to detect errors in a performing ensemble is an essential skill of a music educator (Brand and Burnsed, 1981; Doerksen, 1999; Taebel, 1980). As seen throughout the literature, there has been a great deal of inquiry as to what helps to build and increase these skills. Byo’s (1997) participants had an overall response rate of less than 50%, indicating that error detection is perhaps a weak skill for a majority of musicians. If Byo’s study is any indication of the general population of music educators, then the area of error detection needs to be further explored to determine how to build and refine these skills.

Success in error detection ability has been seen in the implementation of programmed materials. However, DeCarbo (1981, 1982) found that undergraduate students who received additional error detection training in a conducting classroom performed significantly higher on conducting an ensemble and detecting errors than undergraduate students who did not receive the conducting experience but only programmed instruction. This indicates that perhaps learning these skills in a classroom setting will not transfer to conducting on a podium or in front of a classroom of students. Additionally, Forsythe and Woods (1983) determined that students were less successful at performing error detection when participating in the act of conducting, further indicating that testing and instruction may not fully prepare a music educator for detecting errors when leading and conducting an ensemble. In addition, it has been determined that as the number of parts in the musical examples increase, the error detection scores tend to decrease (Byo, 1997; Crowe, 1996; Sheldon, 1998), further supporting the idea that music educators may not be prepared for detecting errors in a large ensemble setting, such as a concert band.

Throughout the error detection programmed materials and tests that have been developed to examine error detection abilities (DeCarbo, 1981, 1982; Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971), perhaps the most commonly used error detection test in music education research to investigate error detection ability amongst music educators, both pre-service and in-service, is the *MLR: Instrumental Score Reading Program* (Doane, 1988; Forseth & Grunow, 1979; Forsythe & Woods, 1983; Hochkeppel, 1993; Van Oyen and Nierman, 1998). However, this test, developed from the *Visual-Aural Discrimination Skills* (Grunow, 1980), does not prove to be the most reliable and internally consistent error detection test. Using split-halves reliability testing twice, the test administration resulted in an overall reliability coefficient of .66 and .71, with the general music criteria resulting at an alpha level of .40 and .34 and the specific technical criteria at .74 and .81. These inconsistent numbers are both above and below the threshold typically used as a reference point when examining the reliability of an instrument in the social sciences (Field, 2009).

The error detection tests discussed above use recordings of two-, three-, four-, or five-part performances (Froseth & Grunow, 1979; Grunow, 1980; Sidnell, 1971), use only brass instruments for the recordings (DeCarbo, 1981, 1982), and perform the music in only concert keys (Byo, 1993; Sidnell, 1971). Ramsey (1978, 1979) created an error detection test that did prove to be internally consistent, with a coefficient of .71 and used full band literature; however, the method in which he chose the literature used for detecting errors was by choosing the error first and then choosing accompanying literature. For example, if he wanted an error that included students missing a key change, he specifically chose repertoire that contained a key change. While this is one method of choosing repertoire, I felt that by choosing the repertoire first and then deciding on the errors that may occur in the musical excerpts the ecological validity would be better supported. In addition, his test was developed to accompany programmed materials to help improve error detection, which were not a part of this present study.

I sought to have an error detection test that would not only help music educators determine errors in a setting that was as similar to an authentic instrumental music performance situation as possible, but also to serve as a valid and reliable tool in determining a music educator's error detection ability. Therefore, the purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances.

Chapter 3: Methodology

Restatement of the Purpose

The purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances.

Test Design

The first step in the development of the *IMEDT* was to examine other musical assessment tools as to help me shape the types of questions and items that would appear on the test. I started by examining other tests that were used in error detection research, as discussed throughout the first two chapters to help frame decisions such as the level of band music to use and the number and types of errors to include in the test. I then explored methodological sources to create a foundation for the test to ensure it was a valid and reliable tool. Boyle and Radocy (1987) discuss criterion for creating a test, which I used to help guide decisions.

As the test designer I needed to decide what was important to be tested (Boyle & Radocy). I did this by examining previous literature on error detection. For example, pitch and rhythm errors are described as "typical errors" encountered during a rehearsal (Byo, 1993; Ramsey, 1978, 1979) and most important to include as errors on the test.

After the test was developed, it was important to examine the difficulty of items with an item analysis to determine the strengths and weaknesses of the test (Boyle & Radocy, 1987). This process was completed in the development of the *IMEDT* following the pilot study. I completed an item analysis to determine if all errors were audible on the recordings and made any needed changes. These item analysis and other methodological steps used in the creation of the *IMEDT* will be discussed in the following sections of this chapter; (a) musical selections; (b) pilot study; (c)scoring the data; (d) development and administration of six test variations; (e) participants; (f) test administration and; (g) *IMEDT* in score and recording with non-controlled time (S&R/N) method.

Musical Selections

Selecting music

The eight musical excerpts used as stimuli for evaluation were sections of five band works, each of which was rated as grade three in the *Teaching Music through Performance in Band* series (Miles, 1997, 2000, 2004). Excerpts were selected that depicted a variety of tonalities (major and non-major), tempi (fast and slow) and meters (simple and non-simple) and defined as follows:

Major: Excerpt contained a tonal center in one of the twelve major scales/keys.

Non-Major: Excerpt contained a tonal center not found in one of the twelve major scales/keys (e.g. minor, Dorian, etc).

Fast: The minimum macro beat for the tempo was 126 beats per minute.

Slow: The maximum macro beat was 84 beats per minute.

Simple meter: The beat divides into two equal parts. This can be simple duple meter (e.g., 2/4 time signature), simple triple meter (e.g., 3/4 time signature), or simple quadruple meter (e.g., 4/4 time signature) (Kostka & Payne, 2009). Simple beat also included musical excerpts that did not change time signatures.

Non-simple meter: The beat cannot be divided into two equal parts. For the study, compound beat, when the beat divides into three equal parts (e.g. 6/8 time signature) or mixed meter, a composer using rapidly changing time signatures (Kostka & Payne, 2009) were used to demonstrate non-simple meter.

The musical excerpts were each 8-12 measures in length and followed a musical phrase that began at the start of a phrase and concluded at the end of a phrase. Using three musical criteria, divided into two possibilities, eight categories were developed as can be seen in Table 1. These categories represented separate combinations of tonality, tempo, and meter. Once these categories were determined, a musical excerpt was chosen to fulfill each of the eight categories. Therefore each musical excerpt represented a different combination of the six musical criteria.

Table 1

Musical Excerpts used on IMEDT based on Musical Criteria

Musical Excerpt	Tonality	Tempo	Meter
American Riversongs (M. 108 - 119) Pierre La Plante	Major	Fast	Simple
American Riversongs (M.1-12) Pierre La Plante	Major	Fast	Non- Simple
American Riversongs (M.82-90) Pierre La Plante	Major	Slow	Simple
The Renaissance Fair (Movement 1, M. 12 – 21) Bob Margolis	Major	Slow	Non- Simple
Rollo Takes a Walk (M. 35 – 45) David Maslanka	Non-Major	Fast	Simple
Highbridge Excursion (Movement 2, M. 84-92) Mark Williams	Non-Major	Fast	Non- Simple
Cajun Folk Songs (Movement 1, M. 17-29) Frank Tichelli	Non-Major	Slow	Simple
Cajun Folk Songs (Movement 1, M. 29-39) Frank Tichelli	Non-Major	Slow	Non- Simple

Error selection and insertion

Once the musical excerpts were selected, a panel of four expert band directors was chosen to examine the excerpts. Goolsby (1996, 1997, 1999) defined an expert band director as having at least six years of teaching experience. This definition was used in the current study. The first expert taught band at the elementary and middle school levels for eleven years. The second expert was a band director at a local community college who had approximately 12 years of middle school and college band directing experience. The third member of the panel was a high school instrumental music teacher in the final stages of completing his doctoral work in music education. The fourth member of the panel was a doctorate student in wind band conducting, with six years of high school teaching experience prior to working on his doctorate.

The panel was the emailed the -musical scores of the eight excerpts in September 2011. I asked them to identify the types of pitch, rhythm, and articulation errors that might be made in a typical rehearsal of the music with a middle school or high school band. They were asked to identify the instrument or instruments where the error would typically occur and to describe the nature of the error. A few examples of responses I received from the panel were:

"Measure 40-Every flute player will miss the A-natural"

"Measure 9-Concert A-natural will be missed"

"Measure 83/84 – Trumpets will hold the tie too long and rush the eighths"

These data were then coded and examined for common trends among the panel.

A second panel comprised of three university music education faculty instructors, including me reviewed the suggested errors and made final decisions as to what errors would be used in the present study. A total of 10 pitch errors, 14 rhythm errors, and 8 articulation errors were selected for inclusion in the instrument.

Scores and parts for each piece were secured and separate parts for each musical instrument in the wind ensemble were created, using the music notation software *Sibelius* (version 3). Scores and instrument parts with errors were identical to the original parts except for the inclusion of the specified errors. The errors were highlighted in the corresponding part to better assist the performer in identification of the errors so they would not be surprised or confused when performing.

The University of Maryland Symphonic Wind Orchestra, a 50-member ensemble comprised of graduate and undergraduate students and conducted by Dr. Michael Votta, performed the excerpts for the study during a regularly scheduled rehearsal in October 2011. This rehearsal was recorded by the recording, editing, and mastering engineer under contract by the University of Maryland in a large professional performing hall. The ensemble was recorded performing several versions of each excerpt's model performance and enough versions of the musical excerpts with the inserted errors until the wind orchestra conductor, the recording engineer, and I were satisfied with the quality and performance of the musical excerpt. The instrumentation used in the ensemble on the recordings can be found in Appendix A.

From the versions recorded by the wind orchestra, I selected a "model performance" and an "altered performance" to be included on the final recording. These 16 recordings were chosen based on the clarity and precision of the musical performance, the volume and performance of the errors. I enlisted a music education faculty member to assist with the examination of the final recordings to ensure that each error was clearly audible. Eight errors were eliminated because the error was inaudible on the recording. Additionally, three performance errors were made during the recording sessions that were not among the intended errors. These were retained for use in the current study as I believed the mistakes made by the university ensemble would also be made by Younger, less experienced musicians. A total of 10 pitch errors, 11 rhythm errors, and 6 articulation errors were used in the final version of *IMEDT*.

Pilot Study

The Instrumental Music Error Detection Test (*IMEDT*) was developed from these recordings. The recordings were uploaded onto iTunes on my laptop computer, a Gateway PC Notebook. Logitech Speaker System Z320 was used to allow participants in the study to clearly hear all instrumental parts in each musical excerpt. Each participant was provided a packet that contained full transposed scores for each of the musical excerpts, with measures numbered consecutively. Participants were informed that each musical excerpt contained two to five errors, which could be pitch, rhythm, or articulation errors.

To create the packet of musical scores, I photocopied only the measures from the score that would be needed to complete the *IMEDT*. In addition, I eliminated all markings of the title of the piece and the composer of the piece as some participants may have been familiar with the band pieces on the *IMEDT* which could have been a threat to validity. All other musical aspects of the piece of music were left intact. I replaced the title of each piece with a letter. These letters were used throughout this current research to represent each musical excerpt. The letters assigned can be seen in Table 2.

Table 2

Letters Used to Represent each Musical Excerpt

Letter	Musical Excerpt
A	*American Riversongs* (M.1-12): Pierre La Plante
B	*American Riversongs* (M.82-90): Pierre La Plante
C	*American Riversongs* (M. 108 - 119): Pierre La Plante
D	*Cajun Folk Songs* (Movement 1, M. 17-29): Frank Tichelli
E	*Cajun Folk Songs* (Movement 1, M. 29-39): Frank Tichelli
F	*Rollo Takes a Walk* (M. 35 – 45)*:* David Maslanka
G	*The Renaissance Fair* (Movement 1, M. 12 – 21): Bob Margolis
H	*Highbridge Excursion* (Movement 2, M. 84-92): Mark Williams

Also contained in the packet was an answer sheet that corresponded to each musical excerpt with columns for participants to indicate the measure, beat, and type of error that occurred, as well as a section where they could describe the nature of the error. Since participants were told that there could be anywhere from two to five errors in each example, I did not want the answer sheet to influence the participants' answers, therefore five spaces were given for each excerpt. I decided to provide participants with an answer sheet rather than writing their answers on the musical score to avoid any unclear or ambiguous circles or writing on the musical score itself that may lead to confusing answers I was not able to decipher. The answer sheet used in the present study can be found in Appendix F.

In addition to taking the *IMEDT*, each participant completed an accompanying questionnaire. This sixteen item questionnaire was designed to collect demographic information about the participant. The questionnaire was created using the online tool SurveyMonkey (surveymonkey.com) and contained questions regarding the music educator's current teaching practice (e. g. How many years have you been teaching in a K-12 setting?), performance experience (e.g. What is your primary instrument?), and educational background (e.g. What is your highest degree earned?). Each participant was sent the link to the questionnaire prior to taking the test. A copy of the questionnaire can be found in Appendix G.

Pilot test data

Pilot data for the study were collected at three separate points to examine the validity and reliability of the *IMEDT*. The instrument was first implemented in a graduate music education class, consisting of seven former or practicing K-12 music educators, a graduate student in music performance who formerly instructed in a K-12 setting, and one music education faculty member.

Participants were given the following directions:

1. Examine the score for one minute.
2. Listen to a model performance of each excerpt in its entirety twice through while following the musical score
3. Listen to the altered version with the inserted errors
4. Make notes as needed on the score or on the answer sheet for 90 seconds
5. Listen to the altered version a second time.
6. Finish completing the answer sheet for thirty more seconds.

A number of adjustments were made following this initial administration of the test. Participants felt that they did not need a full minute to examine the score prior to listening to the model performance; therefore the time was adjusted to 30 seconds. Additionally, they preferred a shorter amount of time to take notes on the score, and more time to complete the answer sheet at the end. Finally, there was some confusion concerning the process for completing the answer sheet, which resulted in a sample answer box being inserted at the top of the answer sheet to serve as a model. These changes were made to the procedures for a second administration of the test a few days later that involved five undergraduate music education majors at the University of Maryland, who were in their third or fourth year of study. These students had completed all music theory and aural skills coursework and were enrolled in or had completed their conducting coursework, which included instrumental score study. Participants completed the following procedures for the second administration:

1. Examine the score for thirty seconds.
2. Listen to the model performance example twice through while following the musical score
3. Listen to the altered version with the inserted errors
4. Make notes as needed on the score or on the answer sheet for one minute.
5. Listen to the altered version a second time.
6. Finish completing the answer sheet for thirty more seconds. Additional time was given to write answers on answer sheet, however, no additional score study time was given.

During the first and second pilot test administrations, the *IMEDT* was given in a university classroom, with all participants sitting around a table. It was clear during these administrations that participants were impacted by the note-taking of their classmates . I observed the participants completing the test and noticed if one person marked the score while the music was playing, this drew attention to that measure in the music, which was then listened to by members of the class with more scrutiny the second time the example was played. This was noticeable in behaviors such as participants looking at their classmates while they circled a measure in the musical score or putting a finger on that measure before the music was played a second time. Therefore, it seemed appropriate that the test would be given individually to eliminate the possibility for participants to influence each other's answers and cause threats to validity.

The test was then administered to three additional expert band directors. I traveled to the school that each participant was employed at to administer the *IMEDT* in a individual setting. Each participant had at least ten years of instrumental music teaching experience in a K-12 setting. The same procedures were used as in the second pilot study, with two small adjustments. During the fourth step, rather than encouraging the participants to not write on the answer sheet, I informed them to wait until the end, after they had heard the altered version of the recording twice. Secondly, after the participants listened to the excerpt twice as a model performance and twice altered (following step six described above), I asked them to turn the score over after one minute to eliminate any additional score study time. These three participants were provided with additional time to fill in the answer sheet if needed; however no additional score study was provided. The procedures for the test administration for these three participants were as follows:

1. Examine the score for thirty seconds.

2. Listen to the model performance example twice through while following on the musical score
3. Listen to the altered version with the inserted errors
4. Make notes as needed on the score for one minute.
5. Listen to the altered version a second time.
6. Finish completing the answer sheet for thirty more seconds. Additional time was given to write answers on answer sheet, however, after thirty seconds have participants hand me the score.

The accompanying questionnaire for the *IMEDT*, the *Demographic Information Form,* was also used during the collection of pilot data. In the first administration of the test with the graduate students in music education, a hard copy of the questionnaire was distributed. Participants took the questionnaire as a group while I was in the room. It was during this time the participants discussed any confusion with the questions or answers. There were two questions which participants found difficult and therefore adjustments were made. For example, there was one question regarding the performance level of the participant's ensemble. In the original questionnaire, the participant would answer how a panel of instrumental music education colleagues would rate his or her "best" concert band on a scale of one to seven using a Likert-type scale. However, the music teachers who had competed in festivals and competitions found this question to be a little confusing and therefore the scale was changed to ratings typically used in music festivals and competitions, superior, excellent, good, fair, and poor, thus ensuring the language of the answer was more friendly and familiar to the participant. However it was decided that these questions were beyond the scope of this study and were eliminated from the *Demographic Information Form*.

The revised *Demographic Information Form* was then distributed as an online tool for the second pilot group. All participants in this group completed the *Demographic Information Form* online and submitted their responses through SurveyMonkey (surveymonkey.com).com. During the administering of the *IMEDT*, I asked participants for any feedback or confusion on the revised questionnaire and they stated that it was clear and easy to follow. The same response was given by the expert band directors who completed the third phase of the pilot study. Thus, no further changes were made.

The item analysis

An item analysis was performed on the *IMEDT*, which was similar to the analysis used by Grunow (1980) to determine the difficulty level of errors for the *Visual-aural Discrimination Test.* A copy of this analysis can be seen in Appendix H. Each error was examined to determine how many of the 17 participants from the pilot study accurately detected the error, followed by a second analysis which determined how many participants accurately determined the nature of the error. After the item analysis was complete, a few adjustments were made to the *IMEDT*. Accenting and not accenting where appropriate was initially considered an articulation error, though based upon the discrepancies in participants' answers I believed the distinction between accented and non-accented notes as heard on the recordings was not a large enough contrast and created a large discrepancy in the item analysis. Some participants found the accents to be acceptably performed when others did not. In addition, I also discussed the performed accents with a member of the wind ensemble who performed on the recordings, a doctorate student in trumpet performance, to inquire if her opinion whether the distinction was great enough on the recordings and she agreed it was not and informed me she felt this during the recording session as well. Therefore accents were removed from the definition of an articulation error for the present study. Also, in grading the nature section of the test, it was difficult to determine if the participant should earn a point or not, due to the subjective nature of performing accents. Thus, they were eliminated from the test so that articulation errors were defined as tonguing and slurring errors only. Also, an additional pitch error and a rhythm error made by the performers during the recording session were discovered by several participants that had not been uncovered by the music education faculty member or me. The two musical excerpts were then examined by a colleague with perfect pitch. I met with the colleague, played the musical excerpts while he examined the score, and asked him to determine what errors had occurred. He indicated all the errors in the two musical excerpts, including the two errors in question. It was decided they were audible in the recordings and included in the tally of errors. Finally, pitch errors that carried from one measure to the next now counted as two errors if both measures were identified by the participant. In the end, 30 total errors were included in the *IMEDT* consisting of 12 pitch errors, 12 rhythm errors, and 6 articulation errors. The nature of all errors are described in Appendix B and can also be found according to type of error in Appendices C, D, and E.

Scoring of the Data

Different methods of scoring pilot data were examined to obtain optimum validity and reliability of the *IMEDT*. A scoring item analysis was undertaken to estimate reliability and validity of scoring the data. Each measure of music was labeled an “item” and there were 82 items on the error section of the *IMEDT*. The *IMEDT* contained one measure in which two errors occurred. An item was considered to be answered correctly if the participant either (a) identified an error in a measure that had an error or (b) identified no error in a measure that had no error. Correct answers received one point. If only one error was identified in a measure with two errors then a score of .5 was given and if both errors were identified a full point was given. In addition, a participant would receive no credit if they indicated an error in a measure where there was none (a misidentified error). Finally, if a participant identified an error correctly, but also listed a misidentified error in the same measure, a score of .5 was given.

This same type of scoring was performed to determine the nature of the error. This item analysis contained only 30 items, one for each error that occurred throughout the musical selections. The participant received a .5 if the instrument or instruments (e.g., alto saxophone two and clarinet two) were identified in the indicated measure and another half point (.5) if the error was described appropriately (e.g., alto saxophone two and clarinet two are playing a concert Eb instead of a concert E natural). Therefore on this section of the *IMEDT*, a participant could receive no score (zero), a half point (.5), or a full point (1) for each item. If the nature of all items was determined, the highest score was 30 making the range of scores for this section 0-30.

All pilot data were entered into SPSS, Version 17.0. Cronbach's Alpha was used to estimate the reliability of test scores. In relation to the first section of the test, the Error Test in which participants determined the measure, beat, and type of error, the Cronbach's alpha was .659. This number was not at an acceptable threshold (.70) typically used as a reference point when examining the reliability of an instrument in the social sciences (Field, 2009). Cronbach's alpha for the nature section of the test, the Nature Descriptive Test, was .86, indicating that this section of the test was estimated to be reliable and internally consistent

Development and Administration of Six Test Variations

After the pilot study and data analysis were completed, my dissertation committee noted that, I had not explored other methods of administering the test, which may have helped to increase the reliability coefficients. Therefore, six methods of test administration variations were developed based on past practice in error detection literature (Byo & Sheldon, 2000; Crowe, 1996; Grunow, 1980; Hochkeppel, 1993; Hopkins, 1991; Sheldon, 2004; Van Oyen & Nierman, 1998). Past practice has used a combination of test administrations that have included time to study the score prior to hearing musical examples, listening to model performances of musical examples, combinations of score study time and listening to model performances of musical examples, and variations of time provided for these activities. The test administration variations that were chosen for the *IMEDT* based on these practices were:

1. Score study with controlled time (S/C)
2. Score study with non-controlled time (S/N)
3. Recording only with controlled time (R/C)
4. Recording only with non-controlled time (R/N)
5. Score study & recording with controlled with (S&R/C)
6. Score study & recording with non-controlled time (S&R/N)

Test administration variations number one and two, were given with time for the participant to study the score; however no model performance of the musical excerpt was played prior to listening to the excerpts with inserted errors. For test variations three and four, participants heard a model performance of the musical excerpt before the error performance was played; however they received no time to study the score prior to listening to the excerpts. For test administration variations five and six, participants received both score study time and were able to hear a model performance of the musical excerpt before the error performance was played. For all six test administration variations, the error performance was played twice for the participant. Additionally, participants were not given specific directions as to when to write on the score and/or when to write on the answer sheet, rather I just informed participants they were welcome to make any notes on the score or answer sheet they needed as the musical excerpt was being administered. Since the test was to be administered in an individual setting, I felt it was important for each participant to feel comfortable in how and when they chose to make notes throughout the process to fully examine their error detection ability, rather than note taking ability. The full directions for each test administration variation can be found in Appendix I.

A rotation matrix was developed to ensure that each musical excerpt was heard in each method of administration. The first rotation started with musical excerpt A being administered in the first test administration variation (S/C), therefore musical excerpt B administered in the second rotation method (S/N). This pattern continued through the six test administration variations. Once the first rotation was complete, it began again with musical excerpt G. This rotation went through three cycles of the musical excerpts. The rotation matrix was then started a second time with musical excerpt A being performed in the second administration variation (S/N), musical excerpt B in the third test administration variation (R/C), and so forth. This design ensured that all musical excerpts were administered once in each test administration variation. This rotation can be seen in the Table 3 below, where the musical excerpts are labeled with letters, and the methods are listed with the abbreviations described above.

Table 3

IMEDT Testing Method Administration Rotations

	First Rotation			Second Rotation		
Musical Excerpt	1	2	3	1	2	3
A	S/C	R/C	S&R/C	S/N	R/N	S&R/N
B	S/N	R/N	S&R/N	R/C	S&R/C	S/C
C	R/C	S&R/C	S/C	R/N	S&R/N	S/N
D	R/N	S&R/N	S/N	S&R/C	S/C	R/C
E	S&R/C	S/C	R/C	S&R/N	S/N	R/N
F	S&R/N	S/N	R/N	S/C	R/C	S&R/C
G	S/C	R/C	S&R/C	S/N	R/N	S&R/N
H	S/N	R/N	S&R/N	R/C	S&R/C	S/C

Note. S/C = score study with controlled time, S/N = score study with non-controlled time, R/C = recording only with controlled time, R/N = recording only with non-controlled time, S&R/C = score study and recording with controlled time, S&R/N = score study and recording with non-controlled time

From this matrix of musical excerpts and rotation of test administrations, six versions of the *IMEDT* were developed. The musical excerpts were entered into the online tool Research Randomizer (Urbaniak & Plous, 2012) to determine which musical excerpt would be played first for each test, therefore ensuring that each of the six tests was randomized according to musical excerpt order and method of administration. Once this was complete, the six test variations of the *IMEDT* were compiled and ready to administer to participants. The method and order of musical excerpts contained in each test can be seen in Table 4.

Table 4

IMEDT Test in Six Variations of Test Method Administration

Variation 1	Variation 3	Variation 5
C – S&R/C	E – S&R/C	A – S&R/N
D – S&R/N	F – S&R/N	B – R/C
E – R/C	G – R/C	C – R/N
F – R/N	H – R/N	D – S/C
G – S/C	A – R/C	E – S/N
H – S/N	B – R/N	F – S&R/C
A – S/C	C – S/C	G – S&R/N
B – SN	D – S/N	H – R/C
Variation 2	Variation 4	Variation 6
B – S&R/N	D – R/C	F – R/C
C – R/C	E – R/N	G – R/N
D – R/N	F – S/C	H – S/C
E – S/C	G – S/N	A – R/N
F – S/N	H – S&R/C	B – S/C
G – S&R/C	A – S/N	C – S/N
H – S&R/N	B – S&R/C	D – S&R/C
A – S&R/C	C- S&R/N	E – S&R/N

Note. The musical excerpt is represented by the letter on the left hand side, with the abbreviation of the test administration variation listed beside the excerpt.

[a] S/C = score study with controlled time, S/N = score study with non-controlled time, R/C = recording only with controlled time, R/N = recording only with non-controlled time, S&R/C = score study and recording with controlled time, S&R/N = score study and recording with non-controlled time

Participants

Participants for this study included pre-service music educators, in-service music educators, and graduate students in a music performance area who had completed at least one semester of instrumental score study as part of their coursework. G Power (3.1.2) (Faul, Erdfelder, Lang, & Buchner, 2007) was used to determine the number of participants needed for this study. In addition, I examined previous error detection literature and test development literature and found the number of participants used to examine reliability and validity of these instruments to be a mean average of 43.5 participants (DeCarbo, 1981, 1982; Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971). At the $\alpha = .05$ level, a sample size of 81 participants was needed. In order to achieve this sample size, a convenience sampling method, with snowball sampling, was used to contact participants in the greater Washington Metropolitan Area, which includes the District of Columbia as well as Maryland and Northern Virginia. Possible participants were contacted via email in December 2011. A follow up email to non-respondents of the first email was sent out in early January 2012.

To recruit participants for the study, I first created a list of possible participants of current K-12 instrumental music teachers with whom I had knowledge. From this list, I contacted teachers I knew from three varying settings: (a) a summer camp hosted by the University of Maryland, where I serve as an administrator; (b) teachers who served as student teaching mentors for the same institution and (c) friends and colleagues who taught in instrumental music settings in the area. I also contacted any pre-service instrumental music educator I was familiar with who had taken at least one semester of conducting. I asked each subject to recommend other instrumental music teachers who might be willing to participate in the study. In addition, two in-service music educators posted information regarding the study on the instrumental music teachers’ listserv, hosted by their school system. These postings lead to several instrumental music educators contacting me to partake in the study.

Data were collected from the initial group of participants (*n* = 62) who completed the *IMEDT* in the six test administration variations for six weeks in December 2011 and January 2012. After each participant was contacted via email, communication was exchanged to set up a meeting time and place. I typically met participants at their place of employment or at the University of Maryland. Of the initial 62 participants, one participant was a student teaching intern (*n* = 1), a student that has completed all required coursework and was currently completing an instrumental music student teaching internship in a public school setting. Thirteen percent of the participants were pre-service music educators (*n* = 8), 8% percent were current graduate music students in areas other than music education (*n* = 5), and 5% percent were current college or university music instructors (*n* = 3). The largest section of the sample, 73%, was comprised of current music educators working in a K-12 setting (*n* = 45).

Of the current in-service music educators, 27% (*n* = 12), were teaching in an elementary setting, 36% (*n* = 16) were teaching in a middle school setting, and 24% (*n* = 11) were teaching in a high school setting. The remaining 13% (*n* = 6) were teaching both elementary and middle school equally. In addition, there was a variety of years of teaching experience between these K-12 music educators. Using the definitions provided by Goolsby (1996, 1997, 1999), 36% (*n* = 16) were novice teachers, having taught in a K-12 setting anywhere between 1 and 5 years, while 64% (*n* = 29) were considered expert teachers having had at least six years of teaching experience in a K-12 setting. The number of years teaching ranged from one to more than 30 years (*M* = 10.9, *SD* = 8.87).

After one test method was determined to be the most internally consistent using Cronbach's Alpha, the same method of participant recruitment was repeated. Emails were sent out to recruit participants in February 2012 and continued through early March. Twenty participants completed the test in the preferred administration method. Of these 20 participants, one was a student teaching intern, another was a university assistant professor in instrumental music education not at the University of Maryland (*n* = 1), six were pre-service instrumental music educators, and the remaining were in-service instrumental music educators. The participants that were working in K-12 settings were teaching instrumental music in elementary schools (*n* = 1), middle schools (*n* = 4), and high schools (*n* = 6). One participant taught an equal amount between middle and high school instrumental music (*n* = 1). Two were novice teachers, while 10 were considered expert teachers. The number of years of teaching experience ranged from two to 30 or more years (*M* = 8.3, *SD* = 9.47).

Test Administration

I contacted participants via email (see Appendix J) and once they agreed to take part in the study, communication was exchanged to agree on a time and location to meet. The participants were in-service music educators, pre-service music educators, student teaching interns, graduate students in music performance areas and college/university music instructors. The only prerequisite to participate in the study was to have completed at least one semester of a course that included instrumental score study. Each participant’s name was entered onto a spreadsheet and systematically assigned to a test administration variation. The test method assigned was rotated in the order that participants agreed to participate. Therefore the first participant received Variation 1, the second participant Variation 2, the third participant Variation 3, and so forth. With 62 participants completing this first section of the study, each test method administration was used a minimum of ten rotations.

Once a meeting time and location was established, a confirmation email was sent to the participant the week before the meeting. This email also contained the link to the online *Demographic Information Form*. This questionnaire needed to be completed before I arrived at our meeting, as it also contained the consent form to participate in the study. If a participant did not complete this questionnaire online, I asked the questions in person when I arrived at our meeting. The questionnaire can be found in Appendix H.

I met with participants at their place of employment or at the University of Maryland for approximately 45 minutes each to complete the *IMEDT*. Since I completed a majority of data collection over a six week time span, I also employed an assistant, a fellow graduate student in music education to administer the test. We met twice to review and practice the test administration procedures. The directions for each of these test administrations can be found in Appendix I.

***IMEDT* in Score and Recording with Non-Controlled Time (S&R/N) Method**

The six variations of the *IMEDT* were administered to 62 participants and the collected data were analyzed using Cronbach's Alpha to determine the most reliable and internally consistent method of test administration. These results will be discussed in detail in chapter four.

The test administration of score and recording with non-controlled time (S&R/N) displayed a higher alpha level and was chosen as the most reliable and internally consistent method of administration for the *IMEDT*. Once this was determined, the process of contacting participants and data collection was performed a second time in the same methodology as described above, with only one minor adjustment. Throughout test administration, participants would ask the tempo of pieces when they did not receive a model performance. Therefore, when the test was compiled into one full test administration method, the approximate tempo for each musical example was added to the answer sheet (e.g., Example 1-Largo, Quarter Note = 56). All participants in this second phase of administering the *IMEDT* received the preferred test method administration with the new answer sheet, which can be found in Appendix K. The *IMEDT* was then administered to 20 additional participants for a total of 82 participants. These participants were pre-service music educators (*n* = 14), student teaching interns (*n* = 2), K-12 music educators (*n* = 57), graduate music students (*n* = 5), or college/university music instructors (*n* = 4). The results of administering the *IMEDT* using a single method will be discussed in detail in chapter four.

Time Table

This study was conducted over a seven month period. Selection of repertoire, the recording session, and the three pilot studies took place in September and October 2011. Data collection started in the beginning of December and lasted through the beginning of March 2012. I traveled throughout the greater Washington Metropolitan Area throughout December 2011 and January, February, and March 2012.

Summary

This chapter discussed the procedures and methods for this current research study. *The Instrumental Music Error Detection Test*, or *IMEDT*, was administered in six variations to determine the most valid and reliable test administration variation for the test. Once this method was determined the test was administered in its entirety to support these findings. The results of these test method administrations will be discussed in detail in chapter four, while the implications and discussion of these results will be discussed in chapter five.

Chapter 4: Results

Restatement of Purpose

The purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances.

Error Score of *IMEDT*

Reliability and internal consistency

Cronbach's Alpha was used to estimate the reliability and internal consistency of each of the six administration variations examined in the present study. The error scores of 62 participants who completed the test in one of six different test administration variations were used in the analysis. Participants received one point for identifying an error in a measure that had an error or identifying no error in a measure that had no error. In order for a participant to receive the full point they needed to list the measure number and beat number the error occurred on as well as the type of error that occurred (pitch, rhythm, or articulation), no half points were given in this section of the *IMEDT*.

I estimated the coefficient alpha level for each individual musical excerpt in each of the six test administration variations was estimated. For each test, musical excerpts A and G were administered using identical methods, as were musical excerpts B and H. For the purpose of examining the reliability coefficient, the measures, or" items," from these excerpts were combined. The same participant completed one test variation that included musical excerpts A and G in the same test administration variation and musical excerpts

B and H in the same test administration variation. These could be combined using the Cronbach's Alpha method because the participant's ability in error detection was internally consistent, and therefore would not impact the combined scores. Thus Cronbach's Alpha was determined for six musical excerpts (A&G, B&H, C, D, E, and F) in the six different test administration variations. The methods were coded as follows:

1. Score study with controlled time (S/C)
2. Score study with non-controlled time (S/N)
3. Recording only with controlled time (R/C)
4. Recording only with non-controlled time (R/N)
5. Score study & recording with controlled time (S&R/C)
6. Score study & recording with non-controlled time (S&R/N)

The estimate for reliability for each excerpt in each test administration variations can be seen in Table 5.

Table 5

Cronbach's Alpha for Musical Excerpts in Six Test Method Variations

Excerpt	S/C	S/N	R/C	R/N	S&R/C	S&R/N
A&G	-.60	-.73	-.35	.30	-.05	.05
B&H	.49	-.01	.59	.55	.61	.40
C	.15	.21	-.84	.12	-.71	.51
D	.61	-.80	.47	-.28	.00	.56
E	.61	-.34	-.84	-.06	.23	-.19
F	.53	-.39	.41	-.76	.63	.79

Note. S/C = score study with controlled time, S/N = score study with non-controlled time, R/C = recording only with controlled time, R/N = recording only with non-controlled time, S&R/C = score study and recording with controlled time, S&R/N = score study and recording with non-controlled time

[a] Musical excerpts A and G were administered using identical methods, as were musical excerpts B and H for each test rotation variation

I compared the coefficient alphas to determine which testing administration method yielded an estimate closest to the .70 threshold recommended for instrument reliability (Field, 2009). For two musical excerpts, the alpha level was highest in the S/C administration method (excerpt D, α = .61 and excerpt E, α = .61) and for two musical excerpts the alpha level was highest in the S&R/N administration method (excerpt C, α = .51 and excerpt H, α = .79). The item level estimates did not provide clear empirical data to suggest that one test administration variation was superior in reliability to the others. Therefore the mean alpha coefficients for all musical excerpts in each test administration variation were calculated. The results are in Table 6.

Table 6

Mean Alpha Scores for Six Test administration Variations

S/C	S/N	R/C	R/N	S&R/C	S&R/N
.28	-.34	-.18	-.02	.12	.35

Note. S/C = score study with controlled time, S/N = score study with non-controlled time, R/C = recording only with controlled time, R/N = recording only with non-controlled time, S&R/C = score study and recording with controlled time, S&R/N = score study and recording with non-controlled time

After mean scores were calculated, it was clear that the test administration variation of score and recording with non-controlled time (S&R/N) was the closest to the reliability threshold, though it was not met. With a small sample of participants completing each musical excerpt in each test administration variation, it was difficult for any of these to achieve a high reliability coefficient. Since this test administration variation obtained the highest internal consistency of the six test administration variations, it was used to administer the *IMEDT*.

Determination of structure of *IMEDT*

After the optimal test administration variation was determined, it was important to examine the data for most reliable and internally consistent order of musical excerpts to use in the *IMEDT*. Again, using Cronbach's Alpha, I estimated the reliability coefficient for each of the six test variations (Variations 1-6), regardless of the test method administration used. The alpha level for each test, using all eight excerpts, eliminating the variable of "test method administration" is seen in Table 7. Based on the obtained coefficient alpha estimates, the order of musical excerpts in Variation 3 would be used for the *IMEDT* in its full administration.

Table 7

Alpha Levels of IMEDT in Six Test Administration Variations

Variation 1	Variation 2	Variation 3	Variation 4	Variation 5	Variation 6
.68	.68	.83	.78	.60	.77

Additionally, I combined all scores of the initial 62 participants to examine the overall reliability coefficient. This allowed me to examine the internal consistency of each test item, without accounting for test administration variation. Since the alpha levels for each test administration were not meeting the reliability coefficient, it was important to examine the test items to inquire if the low internal consistency was due to the musical excerpts or the test method administration. These combined scores also allowed for a larger sample and larger number of test items to examine. The *IMEDT*, eliminating "test administration variation," resulted in an alpha level of .70. Therefore, the *IMEDT* was estimated to be an internally consistent instrument for measuring error detection ability in regards to pitch, rhythm, and articulation errors regardless of test administration.

After the initial data collection of 62 participants, the *IMEDT* was administered to a new group of participants using only the test administration variation of score and recording with non-controlled time, or S&R/N. The S&R/N test administration variation contained the following steps:

1. The participant examines the score for as much time as needed
2. Listens to model performance
3. Listens to performance with errors
4. Makes notes for as much time as needed
5. Listens to performance with errors a second time
6. Participant completes filling in the answer sheet for as much time as needed

The order of musical excerpts used in Variation 3 from the test rotations found in Table 4, which proved to have the highest alpha level of .83, was used in each administration at this step. To avoid any confusion as the test administrator, I kept the same letters to accompany each musical excerpt. The order was as follows:

1. Excerpt E: *Cajun Folk Songs* (Movement 1, M. 29-39): Frank Tichelli
2. Excerpt F: *Rollo Takes a Walk* (M. 35-45): David Maslanka
3. Excerpt G: *The Renaissance Fair* (Movement 1, M. 12-21): Bob Margolis
4. Excerpt H: *Highbridge Excursion* (Movement 2, M. 84-92): Mark Williams

5. Excerpt A: *American Riversongs* (M.1-12): Pierre La Plante
6. Excerpt B: *American Riversongs* (M.82-90): Pierre La Plante
7. Excerpt C: *American Riversongs* (M.108-119): Pierre La Plante
8. Excerpt D: *Cajun Folk Songs* (Movement 1, M. 29-39): Frank Tichelli

All participants followed the order of musical excerpts from Variation 3 and followed the S&R/N method of test administration. The alpha level for the *IMEDT* in its entirety was .72.

Scores for Error Section of *IMEDT*

The *Instrumental Music Error Detection Test* was administered to two different samples in two separate phases. The first phase empirically determined the optimal test administration variation (*n* = 62) and the second was to begin to estimate the overall reliability estimate of the full *IMEDT* (*n* = 20). This total sample (*N* = 82) was comprised of pre-service music educators, a student teaching intern, graduate students in other areas of music performance, college and university instructors, and in-service music educators who worked at the elementary, middle, and high school levels. It was important to examine the scores to determine the difficulty level and validity of the *IMEDT*. This test contained 82 items; therefore the possible range of scores was zero to 82. Descriptive data for overall test scores can be seen in Tables 8 and 9.

Table 8

IMEDT Scores in Relationship to Professional Status of Initial 62 Participants

Professional Status	*n*	Minimum	Maximum	Mean	SD
Pre-Service Educator	8	52.50	64.50	59.31	4.61
Student Teaching Intern	1	55.00	55.00	55.00	NA
Graduate Student	5	55.50	63.50	58.80	3.49
College/University Instructor	3	54.00	66.50	59.67	6.33
K-12 Educator	45	47.00	70.50	60.87	5.70
Total	62	47.00	70.50	60.35	5.34

Table 9

IMEDT Scores in Optimal Test administration variation of Score and Recording with Non Controlled Time (S&R/N)

Professional Status	*n*	Minimum	Maximum	Mean	SD
Pre-Service Educator	5	52.50	62.50	57.20	3.67
Student Teaching Intern	1	65.50	65.50	65.50	NA
College/University Instructor	1	65.50	65.50	65.50	NA
K-12 Educator	13	56.00	73.50	62.23	5.61
Total	20	52.50	73.50	61.30	5.44

Item analysis of Error Section

An item analysis was performed to examine the difficulty level of each item (measure) that contained an error. Table 10 displays the item analysis results divided into three sections; the initial participants that completed the *IMEDT* in the six different methods (*n* = 62), the participants that completed the *IMEDT* in one method (*n* = 20), and all participants (*N* = 82). I examined how many participants correctly identified the measure, beat, and type of error in that measure. It can be inferred that the higher the number the easier the error was to detect. This item analysis was similar to the analysis used by Grunow (1980) to determine the difficulty level of errors for the *Visual-aural Discrimination Test.*

Table 10

Item Analysis of Difficulty of Pitch, Rhythm, and Articulation Errors on IMEDT

Musical Excerpt	Measure number and Type of Error	First Sample of Administration in Six Test administration variations (*n* = 62)	Second Sample in One Test administration variation, S&R/N (*n* = 20)	Total Sample (*N* = 82)
A - *American Riversongs* (M.1-12) Pierre La Plante	1-R	26	9	35
A - *American Riversongs* (M.1-12) Pierre La Plante	9-A	9	6	15
B- *American Riversongs* (M.82-90) Pierre La Plante	82-P	44	17	61
B- *American Riversongs* (M.82-90) Pierre La Plante	83-P	40	11	51
B- *American Riversongs* (M.82-90) Pierre La Plante	84-R	11	2	13
B- *American Riversongs* (M.82-90) Pierre La Plante	86-R	39	17	56
B- *American Riversongs* (M.82-90) Pierre La Plante	88-P	21	10	31

(Continued)

Musical Excerpt	Measure number and Type of Error	First Sample of Administration in Six Test administration variations (*n* = 62)	Second Sample in One Test administration variation, S&R/N (*n* = 20)	Total Sample (*N* = 82)
C - *American Riversongs* (M. 108 - 119) Pierre La Plante	109-P	16	10	26
C - *American Riversongs* (M. 108 - 119) Pierre La Plante	111-R	6	1	7
C - *American Riversongs* (M. 108 - 119) Pierre La Plante	115-P	47	13	60
C - *American Riversongs* (M. 108 - 119) Pierre La Plante	115-A	13	9	22
C - *American Riversongs* (M. 108 - 119) Pierre La Plante	116-P	13	5	18
D – *Cajun Folk Songs* (Movement 1, M. 17-29) Frank Tichelli	18-P	41	18	59
D – *Cajun Folk Songs* (Movement 1, M. 17-29) Frank Tichelli	21-P	51	18	69

(Continued)

Musical Excerpt	Measure number and Type of Error	First Sample of Administration in Six Test administration variations (n = 62)	Second Sample in One Test administration variation, S&R/N (n = 20)	Total Sample (N = 82)
D – *Cajun Folk Songs* (Movement 1, M. 17-29) Frank Tichelli	22-A	2	2	4
D – *Cajun Folk Songs* (Movement 1, M. 17-29) Frank Tichelli	24-R	9	7	16
E - *Cajun Folk Songs* (Movement 1, M. 29-39) Frank Tichelli	30-R	26	12	38
E - *Cajun Folk Songs* (Movement 1, M. 29-39) Frank Tichelli	33-A	3	1	4
E - *Cajun Folk Songs* (Movement 1, M. 29-39) Frank Tichelli	35-P	38	12	50
E - *Cajun Folk Songs* (Movement 1, M. 29-39) Frank Tichelli	36-R	33	9	42
E - *Cajun Folk Songs* (Movement 1, M. 29-39) Frank Tichelli	37-P	34	8	42
F – *Rollo Takes a Walk* (M. 35 – 45) David Maslanka	35-A	8	0	8

(Continued)

Musical Excerpt	Measure number and Type of Error	First Sample of Administration in Six Test administration variations (n = 62)	Second Sample in One Test administration variation, S&R/N (n = 20)	Total Sample (N = 82)
F – *Rollo Takes a Walk* (M. 35 – 45) David Maslanka	41-P	34	8	42
F – *Rollo Takes a Walk* (M. 35 – 45) David Maslanka	42-R	29	11	40
G - *The Renaissance Fair* (Movement 1, M. 12 – 21) Bob Margolis	13-P	25	7	32
G - *The Renaissance Fair* (Movement 1, M. 12 – 21) Bob Margolis	20-R	39	8	47
G - *The Renaissance Fair* (Movement 1, M. 12 – 21) Bob Margolis	21-R	11	8	19
H – *Highbridge Excursion* (Movement 2, M. 84-92) Mark Williams	87-A	17	4	21

(Continued)

Musical Excerpt	Measure number and Type of Error	First Sample of Administration in Six Test administration variations (n = 62)	Second Sample in One Test administration variation, S&R/N (n = 20)	Total Sample (N = 82)
H – *Highbridge Excursion* (Movement 2, M. 84-92) Mark Williams	88-R	44	12	56
H – *Highbridge Excursion* (Movement 2, M. 84-92) Mark Williams	90-R	18	3	21

Note. P = pitch Error, R = rhythm Error, and A = articulation Error.

In examining the item analysis from the Error Section of the *IMEDT*, I examined the data in two phases. I first examined the data with all participants (*N* = 82) and then with only the participants who completed the test in the S&R/N method (*n* = 20). Overall participants correctly identified pitch errors more than rhythm or articulation errors. Table 11 displays the amount of participants who correctly identified each type of error according to the item analysis of errors in Table 10.

Table 11

Mean and Standard Deviations of Types of Errors Correctly Identified

	Mean Number of Participants that Correctly Identified Errors (*N* = 82)	Mean Number of Participants that Correctly Identified Errors (*n* = 20)
Pitch	45.1 (*SD* = 15.9)	11.4 (*SD* = 4.4)
Rhythm	32.5 (*SD* = 16.8)	8.3 (*SD* = 4.6)
Articulation	12.3 (*SD* = 8.2)	3.7 (*SD* = 3.4)

Nature Descriptive Score of *IMEDT*

Reliability and internal consistency

In order to estimate the reliability and internal consistency of each test administration variation for the Nature Descriptive Section of the *IMEDT*, the section of the test where participants were asked to specify the nature of the error that was performed, the estimated reliability coefficient using Cronbach's Alpha was examined. This initial reliability coefficient was determined after data were collected from the 62 participants who completed the test in the six different test method administrations. In scoring these data, the participant received a .5 if the instrument or instruments (e.g., alto saxophone two and clarinet two) were identified in the indicated measure and another half point (.5) if the error was described appropriately (e.g., alto saxophone two and clarinet two are playing a concert Eb instead of a concert E natural). Therefore, on this section of the *IMEDT*, a participant could receive a score of zero (0), a half point (.5), or a full point (1) for each item. Participants who correctly identified the nature of all errors earned 30 points, making the range of scores on the nature section 0-30.

Due to the limited number of items (n = 30) represented in the Nature Descriptive Section of the *IMEDT*, data for this analysis was pooled across all test methods. When separated by musical excerpts, it was a small number of items to examine for internal consistency. For example, Musical Excerpt A contained only two errors. With approximately ten participants completing each musical excerpt in each test administration variation there was a total of twenty items. To have a higher representation of errors the data were examined after the sample of participants completed the *IMEDT* and again after the second sample. These data can be found in Table 12.

Table 12

Alpha Levels for Nature Descriptive Section of the IMEDT

Test administration variation	*n*	Cronbach's Alpha
All Six Methods	62	.76
Score/Recording & Non Controlled Time	20	.77

Scores for Nature Descriptive Section

The Nature Descriptive Section of the *IMEDT* contained 30 items; therefore, possible scores ranged from zero to 30. The scores for this section of the test were considerably lower than scores from the error section of the *IMEDT*. For the Nature Descriptive Section of the test, the mean score for the initial 62 participants was 8.07 (*SD* = 4.09) while the mean score for the participants who completed the test in the final method of S&R/N procedure was 7.58 (*SD* = 3.85). The test scores for the Nature Descriptive Section of the *IMEDT* can be seen in Table 13.

Table 13

Scores on Nature Section of IMEDT

Administration Method	*n*	Minimum	Maximum	Mean	*SD*
All Six Methods	62	0	16.5	8.07	4.09
S&R/N	20	2	17.0	7.58	3.85
Overall	85	0	17.0	7.95	4.01

Item analysis of Nature Descriptive Section

For each item in the item analysis of the Nature Descriptive Section of the *IMEDT*, a participant could receive no score (zero), a half point (.5), or a full point (1). I examined each item in reference to how many participants earned how many points. This item analysis was similar to the analysis used by Grunow (1980) to determine the difficulty level of errors for the *Visual-aural Discrimination Test.*

For example, for the first error heard in Musical Excerpt A, 19 participants earned a full point, indicating they could identify the instruments that performed the error as well as the error performed. Three participants earned a half point, indicating they could identify either the instruments or the error, but not both. Forty participants earned zero points, indicating they did not correctly identify either the instruments which performed the error or the nature of the error. Scores for both samples are shown in Tables 14 and 15. Higher numbers in the third column, participants who received zero points, indicate the more difficult items.

Table 14

Item Analysis of Nature Section of IMEDT for participants in Six Methods (n=62)

		Number of Participants who received		
Musical Excerpt	Measure Number and Type of Error	0 Points	.5 Points	1 Point
A - *American Riversongs* (M.1-12): La Plante	1-R	40	3	19
A - *American Riversongs* (M.1-12): La Plante	9-A	56	0	6
B- *American Riversongs* (M.82-90): La Plante	82-P	19	8	35
B- *American Riversongs* (M.82-90): La Plante	83-P	22	5	35
B- *American Riversongs* (M.82-90): La Plante	84-R	53	5	4
B- *American Riversongs* (M.82-90): La Plante	86-R	16	11	25
B- *American Riversongs* (M.82-90): La Plante	88-P	49	9	4
C - *American Riversongs* (M.108-119): La Plante	109-P	48	0	14

(Continued)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received: 0 Points	.5 Points	1 Point
C - *American Riversongs* (M.108-119): La Plante	111-R	56	0	6
C - *American Riversongs* (M. 108-119): La Plante	115-P	27	14	21
C - *American Riversongs*(M. 108-119): La Plante	115-A	52	1	9
C - *American Riversongs* (M. 108-119): La Plante	116-P	53	5	4
D – *Cajun Folk Songs* (Movement 1, M. 17-29): Tichelli	18-P	37	29	8
D – *Cajun Folk Songs* (Movement 1, M. 17-29): Tichelli	21-P	17	32	13
D – *Cajun Folk Songs* (Movement 1, M. 17-29); Tichelli	22-A	58	2	2
D – *Cajun Folk Songs* (Movement 1, M. 17-29): Tichelli	24-R	52	2	8
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	30-R	41	10	11
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	33-A	59	0	3

(Continued)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received 0 Points	.5 Points	1 Point
E - *Cajun Folk Songs* (Movement 1, M. 29-39)*:* Tichelli	35-P	55	18	11
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	36-R	45	6	11
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	37-P	45	7	10
F – *Rollo Takes a Walk* (M. 35 – 45) :Maslanka	35-A	55	0	7
F – *Rollo Takes a Walk* (M. 35 – 45) :Maslanka	41-P	38	9	15
F – *Rollo Takes a Walk* (M. 35 – 45) :Maslanka	42-R	37	4	21

(Continued)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received		
		0 Points	.5 Points	1 Point
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	13-P	53	6	3
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	20-R	26	3	33
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	21-R	52	2	8
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	87-A	49	1	13
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	88-R	24	4	34
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	90-R	48	1	13

Table 15

Item Analysis of Nature Section of IMEDT for participants in S&R/N Method (n=20)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received		
		0 Points	.5 Points	1 Point
A - *American Riversongs* (M.1-12): La Plante	1-R	13	4	3
A - *American Riversongs* (M.1-12): La Plante	9-A	14	2	4
B- *American Riversongs* (M.82-90): La Plante	82-P	5	3	12
B- *American Riversongs* (M.82-90): La Plante	83-P	9	1	10
B- *American Riversongs* (M.82-90): La Plante	84-R	19	1	0
B- *American Riversongs* (M.82-90): La Plante	86-R	9	9	8
B- *American Riversongs* (M.82-90): La Plante	88-P	16	3	1
C - *American Riversongs* (M.108-119): La Plante	109-P	13	1	6
C - *American Riversongs* (M.108-119): La Plante	111-R	20	0	0
C - *American Riversongs* (M. 108-119): La Plante	115-P	12	4	4

(Continued)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received 0 Points	.5 Points	1 Point
C - *American Riversongs*(M. 108-119): La Plante	115-A	14	4	2
C - *American Riversongs* (M. 108-119): La Plante	116-P	17	3	0
D – *Cajun Folk Songs* (Movement 1, M. 17-29)*:* Tichelli	18-P	9	7	4
D – *Cajun Folk Songs* (Movement 1, M. 17-29)*:* Tichelli	21-P	8	8	4
D – *Cajun Folk Songs* (Movement 1, M. 17-29); Tichelli	22-A	19	0	1
D – *Cajun Folk Songs* (Movement 1, M. 17-29)*:* Tichelli	24-R	14	1	5
E - *Cajun Folk Songs* (Movement 1, M. 29-39)*:* Tichelli	30-R	9	8	3
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	33-A	19	1	0
E - *Cajun Folk Songs* (Movement 1, M. 29-39)*:* Tichelli	35-P	11	6	3

(Continued)

Musical Excerpt	Measure Number and Type of Error	Number of Participants who received: 0 Points	.5 Points	1 Point
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	36-R	14	0	6
E - *Cajun Folk Songs* (Movement 1, M. 29-39): Tichelli	37-P	16	2	2
F – *Rollo Takes a Walk* (M. 35 – 45) :Maslanka	35-A	20	0	0
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	13-P	20	0	0
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	20-R	12	1	7
G - *The Renaissance Fair* (Movement 1, M. 12 – 21): Margolis	21-R	12	1	7
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	87-A	17	1	2
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	88-R	9	2	9
H – *Highbridge Excursion* (Movement 2, M. 84-92): Williams	90-R	17	1	2

Validity of *IMEDT*

The validity of the *IMEDT* was examined in relationship to four areas of validity; content, criterion-related, construct, and ecological. Table 16 contains the steps taken to support each type of validity.

Table 16

Steps Taken to Examine Validity of IMEDT

Order Steps for Validity were Taken Throughout Present Study	Type of Validity	Steps Taken to Support Validity	Other Research who used Similar Idea
1	Content	Types of rehearsal errors chosen to include in present study	Byo, 1993; Ramsey, 1978, 1979; Sheldon, 2004
2	Content	Musical excerpts sent to panel of experts	Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971
3	Ecological	Used full ensemble for recordings	Ramsey, 1978, 1979
4	Criterion-related	Examined correlations between *IMEDT* scores and aural skills/music theory grades	Brand & Burnsed, 1981; Sheldon, 1998

Content validity

The steps taken in designing and implementing the *IMEDT*, guided by the extant research literature, helped to assist in confirming its content validity. For example, the errors were limited to pitch, rhythm, and articulation, which are all considered specific technical errors (Froseth & Grunow, 1979). Any errors that were considered to be from the "general music criteria," defined by Grunow (1980) as tempo, balance, style of articulation, tone quality, and intonation, were not measured using the *IMEDT*. Grunow stated that these areas, which tested at low reliability coefficients, were too difficult to determine as acceptable or not acceptable because the nature of the material is too subjective. Also, Sheldon (2004) discovered that pitch, rhythm, and articulation errors are better able to be correctly identified than tempo, dynamics/balance, and intonation errors, which are included in Grunow's definition of "general music criteria." Therefore, musical ideas that would be included in the "general music criteria" as defined by Grunow, were eliminated from the *IMEDT*. Additionally, the errors that were included for the development of the *IMEDT* were decided upon by a panel of expert band directors and approved by another panel of music education instructors at the University of Maryland.

Criterion-related validity

Criterion-related validity examines whether scores from an instrument are a good predictor of a specific outcome. This can be completed by comparing the test scores with one or more external criteria (Creswell, 2005; Messick, 1989). As a first step to establish criterion-related validity, I examined the correlation between participant's scores on the *IMEDT* and the question on the *Demographic Information Form* that was distributed, which asked participants to report their overall aural skills/theory grades during their undergraduate career. There was not a significant correlation, [$r(18) = -.015$, $p > .05$]. This was a limited attempt at examining criterion-related validity, for all aural skills grades were self-reported and there is no evidence to support participants were truthful. Furthermore, error detection may not be a skill that is derived or transferred from other aural skills training, such as ear training (Sheldon, 1998). The area of criterion-related validity should continue to be examined for the *IMEDT*.

Construct validity

Construct validity typically examines if measures show stability across methodologies. Furthermore, it inquires if the data reflect true scores or are scores created by the instrument used (Strub, 1989). If constructs are valid according to this description, it can be expected that correlations between measures of the same construct using different methods would be high (Campbell & Fiske, 1959). Throughout the present study, the research questions were not designed to examine correlations between different methodologies in obtaining individual participant's error detection scores. The present study was designed to examine the properties of the test rather than the properties of the individual participants completing the test. Therefore, the data did not allow for the types of correlations discussed by Campbell & Fiske (1959) to be examined.

However, the field of music education has already defined error detection as a measurable skill and construct. Since the literature reviewed in chapters one and two (Blocher, 1986; Brand & Burnsed, 1981; Byo, 1993, 1997; Byo & Sheldon, 2000; Crowe, 1996; DeCarbo, 1981, 1982; Doane, 1988/1989, Doerksen, 1999; Forsythe & Woods, 1983; Froseth & Grunow, 1979; Gonzo, 1971; Grunow, 1980; Hochkeppel, 1993; Hopkins, 1991; Jennings, 1988; Killian, 1991; Larson, 1977; Locy, 1996; Malone, 1985; Ramsey, 1978, 1979; Sheldon, 1998, 2004; Sidnell, 1971; Stuart, 1979; Taebel, 1980; Thornton, 1998; Van Oyen & Nierman, 1998; Waggoner, 2011) clearly defined error detection ability, it was not the primary focus of this present study to examine the constructs and scores of error detection ability in different methodologies.

Ecological validity

Snow (1974) defines ecological validity as "the extent to which the habits or situations compared in an experiment are representative of the population of situations to which the investigator wishes to generalize" (p. 272). This type of validity perhaps took precedence over the content, criterion-related, and constructs validity of the *Instrumental Music Error Detection Test.* I sought to design an error detection instrument that would simulate an authentic instrumental performance setting. The *IMEDT* contained recordings of an ensemble performing in a rehearsal setting, therefore similar to the situation of a typical band rehearsal.

Chapter 5: Conclusions

Purpose Statement

The purpose of this study was to develop an instrument to measure error detection skills among instrumental music educators and to determine the most valid and reliable method in which to administer the instrument. This instrument, *The Instrumental Music Error Detection Test*, or *IMEDT*, examined error detection ability regarding pitch, rhythm, and articulation errors in recordings of wind ensemble performances.

Research Questions

This research sought to investigate the following questions:

1. What is the validity of the *Instrumental Music Error Detection Test (IMEDT)*?
2. What is the reliability of the *Instrumental Music Error Detection Test (IMEDT)*?
3. What is the optimal method for administration of *the Instrumental Music Error Detection Test (IMEDT)*?

Summary

The *IMEDT* was developed to examine error detection ability regarding pitch, rhythm, and articulation errors in recordings of a full wind ensemble. This test was designed to simulate an authentic rehearsal situation. The musical excerpts were selected from grade-three band literature and performed with full instrumentation. These musical excerpts were examined by two panels of experts. The first panel was comprised of four current or former secondary instrumental music educators and the second panel was comprised of three university music education instructors. The musical scores were examined for errors that would most commonly occur in an instrumental rehearsal in regards to pitch, rhythm, and articulation. Once the errors were identified, a university wind ensemble recorded the excerpts. After several attempts, two acceptable recordings were chosen to be included for the present study for each musical excerpt. The completed *IMEDT* contained two recordings of each of the eight musical excerpts, the first as written and the second with inserted errors, for a total of 16 recordings. A total of 30 errors was inserted into the recordings, 12 pitch errors, 12 rhythm errors, and 8 articulation errors.

The *IMEDT* was administered in six different test administration variations to determine the method that was most valid and reliable and had the highest internal consistency. The six variations of test administration for the *IMEDT* were; (a) score study with controlled time (S/C); (b) score study with non-controlled time (S/N); (c) recording only with controlled time (R/C); (d) recording only with non-controlled time (R/N); (e) score study and recording with controlled time (S&R/C) and; (f) score study and recording with non-controlled time (S&R/N). Each test was administered in a individual setting with the participant and either my assistant or me, and took approximately 45 minutes to an hour to complete. Sixty two participants completed this first phase of test administration. Using Cronbach's alpha to estimate reliability and internal consistency, it was empirically decided that the test administration variation of score and recording with non-controlled time (S&R/N) had the highest alpha level. The order of musical excerpts was also determined empirically through this statistical test. Twenty additional participants completed the second phase of test administration variations of the *IMEDT* in the S&R/N method, again in a individual setting, taking approximately 40-45 minutes to complete. After data collection was complete, it was determined that the *IMEDT* was both reliable and internally consistent ($\alpha = .72$).

Conclusions

Validity

Content validity

Content validity was established for the *IMEDT* during the development of the instrument through two procedures; the choice of errors selected and the examination of the errors by expert instrumental music teachers. The *IMEDT* contained three types of errors, pitch, rhythm, and articulation. Any errors that were considered to be from the "general music criteria," defined by Grunow (1980) as tempo, balance, style of articulation, tone quality, and intonation, were not measured using the *IMEDT*. Grunow stated that these areas, which tested at low reliability coefficients, were too difficult to determine as acceptable or not acceptable because the nature of the material is too subjective, therefore they were eliminated from the *IMEDT*. Second, all errors that were inserted were decided upon by a panel of four expert band directors and confirmed by three university instrumental music educators.

Criterion-related validity

As a first step to establish criterion-related validity, I examined the correlation between participant's scores on the *IMEDT* and the question on the *Demographic Information Form* that was distributed, which asked participants to report their overall aural skills/theory grades during their undergraduate career. There was not a significant correlation. However, this lack of correlation may not be attributed to the criterion-validity of the *IMEDT*. For example, all aural skills grades were self-reported and there is no evidence to support participants were truthful. In addition, past research has established that there is not a correlation between aural skills grades and error detection ability (Brand & Burnsed, 1981), indicating that perhaps skills obtained in an aural skills setting may not transfer to error detection ability (Sheldon, 1998). The area of criterion-related validity should continue to be examined for the IMEDT.

Construct validity

The present study sought to examine research questions pertaining to the development of the *IMEDT* and focused on the properties of the tests rather than properties of the participants completing the test. This type of data collection did not allow for any types of correlations to be examined that might further support the idea of construct validity (Campbell & Fiske, 1959). Future research could examine the scores of participants on the *IMEDT* in correlation with a variety of variables, including other error detection tests, such as the *MLR: Instrumental Score Reading Program* (Froseth & Grunow, 1979) to further support construct validity of the *IMEDT*.

Ecological validity

Snow (1974) defines ecological validity as "the extent to which the habits or situations compared in an experiment are representative of the population of situations to which the investigator wishes to generalize" (p. 272). This type of validity perhaps took precedence over the content, criterion-related, and constructs validity of the *IMEDT*. I sought to design an error detection instrument that would simulate an authentic instrumental performance setting. The *IMEDT* contained recordings of an ensemble performing in a rehearsal setting, therefore similar to the situation of a typical band rehearsal that would take place during a K-12 instrumental music educator's day at work.

Reliability

The *IMEDT* was comprised of two sections. The first section, the Error Test, required participants to identify the measure and beat on which the error occurred as well as the type of error that occurred, either a pitch error, rhythm error, or articulation error. Using Cronbach's alpha, the reliability coefficient was determined for each musical excerpt in each of the six test administration variations. The alpha level of the Error Section of the *IMEDT* was examined after the initial phase of participants (n = 62), and was found to meet the reliability threshold, regardless of test administration variation. The alpha level was examined a second time for the participants who completed the *IMEDT* in the method of score and recordings with no time constraints, S&R/N, (n = 20) and the alpha level was found to be .72. This section of the *IMEDT* proved to be above reliability threshold.

The second section of the *IMEDT* required participants to identify the instrument or instruments that performed the error (e.g., trumpet parts 2 and 3) and to appropriately describe the error. The alpha level of the nature section of the *IMEDT* was examined after the initial group of participants (n = 62), and was estimated to be at .76, regardless of test administration variation. The alpha level was examined a second time for the participants who completed the *IMEDT* in its chosen method of score and recordings with no time constraints, S&R/N, (n = 20) and the alpha level was estimated to be .77.

Optimal Method of Test Administration of *IMEDT*

Cronbach's alpha was used to estimate the reliability coefficient and internal consistency of each of the six test administration variations in this present study. To estimate the reliability coefficient, I first examined the Error Section of the *IMEDT* from the initial 62 participants who completed the test in the six different test administration variations. The alpha level for each individual musical excerpt in each of the six test administration variations was examined The musical excerpts in the score and recording with non-controlled time (S&R/N) had the highest alpha level of the six test administration variations.

After the test administration variation was determined, the most reliable and internally consistent order of musical excerpts was also examined. Using Cronbach's Alpha, I calculated the alpha level for each variation (Variations 1-6), regardless of the test administration. Variation 3, where the order of the musical excerpts was E, F, G, H, A, B, C, and D, displayed the highest alpha level ($\alpha = .83$) and was therefore used as the order of musical excerpts that would appear on the *IMEDT* in the method of S&R/N.

I noticed, as an observer who administered the test that participants seemed to find musical excerpts at faster tempos more difficult. Variation 3 had musical excerpts A and H in the middle of the test, which were the musical excerpts with the fastest tempos on the *IMEDT*. This may account for this variation being the most internally consistent because with this order of musical excerpts, participants had three excerpts in slower tempos prior to excerpts A and H. This allowed participants to adjust and become comfortable with the procedures of the test prior to the fastest excerpts, which lead them to complete these excerpts with the same error detection skill level in which they completed the other excerpts.

The completed version of the *IMEDT* was administered in an individual setting with each individual participant. This process of test administration took approximately 40-45 minutes to complete. Since there was no time constraints given on the test administration, test taking times would vary, however, as the test administrator, I observed it was approximately a 40-45 minute process to complete the test. In addition, the test was administered primarily in the music educator's place of employment (e.g., the band room, the educator's office); however, some were distracted by this. It was not uncommon during these meetings for students to interrupt the educator with a question about another topic, bells to ring, or fellow colleagues or administrators to enter the room. A few participants chose to come meet me at the University of Maryland campus and therefore I reserved a classroom for us to meet in. I felt this was the optimal method of administration, since there were no interruptions and participants could remain focused on the *IMEDT*.

Discussion

A music educator must continuously recognize different types of errors while students perform in order to assist with student improvement during rehearsals and performances (DeCarbo, 1982). Educators who have developed accuracy in error detection are able to identify and label problems in a musical performance, which is a fundamental component of teaching and learning in a music classroom (Sheldon, 2004). However, when pre-service music educators receive aural skills training, it is often focused on melodic and harmonic dictation and sight singing and has no component devoted to the development of error detection skills (Larson, 1977; Sidnell, 1971). Sidnell stated that there was no instructional method for the teaching of score-reading skill. He defined score reading as "the skill in the detection and identification of pitch and rhythm errors in instrumental performance" (p. 85). It appears that 40 years after this statement was made, that it may still hold true.

Error detection training programs and testing materials for instrumental music have been developed over the past 40 years (DeCarbo, 1981, 1982; Froseth & Grunow, 1979; Grunow, 1980; Ramsey, 1978, 1979; Sidnell, 1971), however there is no indication the extent to which these materials are currently used in music teacher education training. These programs have proven to be successful in improving error detection ability in pre-service music educators (DeCarbo, 1981; Doane, 1988; Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971). However, DeCarbo (1981) stated it is unknown if these programs are currently being implemented in music education training programs or if training in these programs will transfer to skills needed in a classroom a setting.

Scholars have worked to try to determine what abilities, experiences, and backgrounds may impact a music educator's error detection ability; however, there is inconclusive evidence in regards to the experiences and backgrounds that may impact the development of this skill, including teaching experience (Byo, 1993; Crowe, 1996; Doerksen, 1999; Forsythe & Woods, 1983; Gonzo, 1971; Grunow, 1980), grade level taught (Grunow, 1980), performance experience (Brand & Burnsed, 1981; Thornton, 1998), music theory and/or aural skills grades (Brand & Burnsed, 1981; Fiske, 1977; Larson, 1977), private lesson experience (Brand & Burnsed, 1981), highest degree earned (Gonzo, 1971; Grunow, 1980; Hewitt, 2007), and primary performance instrument (Blocher, 1993; Hewitt, 2007; Hewitt & Smith, 2004). Before the experiences and backgrounds of music educators can be examined to try and determine their ability to detect errors in musical scores and performances, it is important to have a valid and reliable instrument to measure this skill. I sought to develop a valid and reliable test for instrumental music educators that would measure error detection ability.

Through an item analysis of the Error Test of the *IMEDT* it was noticeable that more participants correctly identified pitch errors than rhythm errors. Scholars have discovered that participants found rhythm errors with greater accuracy than pitch errors (Byo, 1993; Sheldon, 1998). However, the tests used by these scholars were created with a digital keyboard or MIDI synthesizer. This present study is in contrast to this prior research, however, it does reinforce the conclusions discovered by Waggoner (2011) that pitch errors are more correctly identified in a full ensemble texture setting over rhythm errors.

This current research study did not use an experimental or correlational design; therefore there were no specific threats to validity. However, when examining the reliability, internal consistency, and validity of the *IMEDT*, there are some items that should be taken into consideration. First, it is unknown whether a participant may have been familiar with a particular musical excerpt that was used in the study. A participant's familiarity with a musical excerpt may have impacted the data. If a participant had low error detection ability, but was familiar with a particular musical excerpt, this may increase their score for that musical excerpt and therefore impact the internal consistency score. I attempted to control for this by not informing participants of the names of the pieces and removing all titles and composers from the written scores.

The area of criterion-related validity may also need further examination. Criterion-related validity examines whether scores from an instrument are a good predictor of a specific outcome. This can be completed by comparing the test scores with one or more external criteria (Creswell, 2005; Messick, 1989). The purpose of this study was to develop an instrument to measure error detection skills among pre-service and in-service instrumental music educators. The scores obtained from testing were not used to predict a specific outcome. Future research may investigate the correlation between scores on the *IMEDT* and relationships between error detection skills and a specific outcome, such as the participant's ensembles' scores at a festival or contest, the musical aptitude of the participant's performing ensemble, or the participant's score on another error detection test, such as the *MLR: Instrumental Score Reading Program* (Froseth & Grunow, 1979) to help support the criterion-validity of this instrument.

The nature of this study was to create an authentic rehearsal situation using a live wind ensemble; however, with a large performing ensemble, such as the one used in the present study, there are problems that can occur during a recording session. For example, there were some balance discrepancies that were heard on the recordings, however were not audible to the conductor or myself during the recording session. For example, the piccolo part is quite prevalent over other instruments of the ensemble in a few of the recordings. While this will not directly impact the results of the *IMEDT*, it can be distracting to a participant. Although electronic created recordings using music software would have eliminated these discrepancies, I wanted to achieve ecological validity by using a large ensemble; therefore these discrepancies were accepted as part of simulating an authentic rehearsal setting.

Implications for Music Education

It is generally agreed that error detection is an imperative skill for conductors (Byo, 1997; DeCarbo, 1982; Doane, 1989; Forsythe & Woods, 1983; Grunow, 1980; Hochkeppel, 1993; Ramsey, 1978, 1979; Sheldon, 1998; Taebel, 1980); however not only is there not an agreed method of teaching this skill, it is still unclear of this skill can be taught and if so, where it belongs in the curriculum. It has been suggested that the primary way to advance error detection skills is for pre-service music educators to practice on a podium in front of an ensemble (DeCarbo, 1982; Doane, 1989); however it has also been established that the physical act of conducting may have little effect on or even impair error detection ability (Blocher, 1986; Forsythe & Woods, 1983). Ramsey (1978, 1979) found that programmed instruction using full-score band literature seems to be a viable means for training college music students in the skill of detecting pitch and rhythm errors; however, DeCarbo (1982) found that success in programmed instruction does not transfer to error detection ability in front of a live ensemble.

Programmed instruction in error detection can create positive results (Doane, 1988; Malone, 1985; Ramsey, 1978, 1979; Sidnell, 1971). In addition, programmed instruction using full-score band literature may be the viable method for training pre-service educators in error detection (Ramsey, 1978, 1979). Therefore, by using the procedures and methods that were implemented into the development of the *IMEDT*, accompanying programmed materials could also be developed. These programmed materials with full-band literature could be used in a setting such as an instrumental music education methods course to prepare students for detecting errors in front of an ensemble.

An alternative to programmed instruction is to use the material from the *IMEDT* as an assessment tool in a lab band setting. It has been suggested that the primary way to advance error detection skills is for pre-service music educators to practice on a podium in front of an ensemble (DeCarbo, 1982; Doane, 1989), therefore by using the written musical excerpts with the inserted errors, pre-service music educators could practice error detection with their peers. Since skills obtained in an aural skills setting may not transfer to error detection ability (Sheldon, 1998), programmed materials or implementing the musical excerpts from the *IMEDT* into a lab band setting could benefit music teacher education programs.

An efficient and effective rehearsal, as well as a musically expressive performance, may be dependent upon the music teacher's error detection ability, in both speed and accuracy (Brand and Burnsed, 1981; Byo, 1993; Crowe, 1996). The *IMEDT* may serve as an assessment tool in examining this ability of students who have completed their music teacher education coursework and are preparing to enter student teaching internships. Scores obtained using the *IMEDT* can inform student teaching interns as well as their mentor teachers and university supervisors of any weaknesses in error detection. Since increased practice in front of an ensemble enhances error detection ability (DeCarbo, 1982; Doane, 1989), student teaching interns could focus on practicing and improving this skill during the student teaching experience, prior to the student teacher entering the field of music education. By working on these skills prior to graduation, perhaps music educators in their first year of K-12 teaching will be better prepared to perceive and diagnose errors among instrumental ensembles.

Recommendations for Future Research

Research in the area of error detection appears to show conflicting results as to what experiences and backgrounds may impact a person's error detection ability. For example, Crowe (1996) stated that experience was the only factor he investigated that was related to error detection ability. Others have determined that there was no significant difference in error detection scores between music students at the undergraduate level and music students at the graduate level, indicating that experience alone may not impact this skill (Byo, 1993; Forsythe & Woods, 1983; Grunow, 1980). Researchers should further explore what experiences and backgrounds impact error detection ability by using the *IMEDT*.

By using the *IMEDT*, future investigation could also focus on examining correlations between error detection scores and variables that have produced mixed results in music education research, such as; teaching experience (Byo, 1993; Crowe, 1996; Doerksen, 1999; Forsythe & Woods, 1983; Gonzo, 1971; Grunow, 1980), grade level taught (Grunow, 1980), ensemble and performance experience (Brand & Burnsed, 1981; Thornton, 1998), music theory and/or aural skills grades (Brand & Burnsed, 1981; Fiske, 1977; Larson, 1977), private lesson experience (Brand & Burnsed, 1981), highest degree earned (Gonzo, 1971; Grunow, 1980; Hewitt, 2007), and primary performance instrument (Blocher, 1993; Hewitt, 2007; Hewitt & Smith, 2004). Grunow (1980) suggested that research is needed to examine the relationship between visual-aural discrimination skills and the success of an instrumental music educator. He states this can be done through student, peer, and administrative evaluations. Certainly, this is an area of future research that should be explored using the *IMEDT*.

Another area in which the *IMEDT* may serve as a beneficial assessment tool for music education research is the examination of the types of errors music educators are successful in detecting and the musical setting in which the errors take place. Scholars have discovered that participants found rhythm errors with greater accuracy than pitch errors (Byo, 1993; Sheldon, 1998) and in contrast, that in a full wind ensemble setting, participants had greater accuracy in determining pitch errors as opposed to rhythm errors

(Waggoner, 2011), as was also found in this present study. In addition, by examining the item analysis performed on the errors of the *IMEDT,* I have observed that the more difficult errors to detect are in the middle voices of the ensemble (e.g., French horns). This relates to Sheldon's (2004) finding that pre-service music educators labeled errors in voice one, the soprano voice, more accurately than errors in the lower voices. The *IMEDT* may be an effective research tool in examining these discrepancies of what types of errors music educators are most successful in detecting and to what degree the voice placement of errors impacts error detection ability.

In a survey completed by 173 undergraduate conductors, Silvey (2010) found that students were least confident with error detection and correction when conducting or rehearsing an ensemble. This current study sought to develop a test that could measure error detection ability, however with participants in this study, or in the study conducted by Silvey, how can music teacher educators help students who are not as successful as error detection? This current study could perhaps lead to the next step that needs to be taken in music education research relating to error detection; now that we know a pre-service educator is weak in this area of error detection, what can music teacher educators do to help it improve?

The *Instrumental Music Error Detection Test* was developed to examine error detection ability in pre-service and in-service instrumental music educators. This instrument was developed with the intentions for use by music education practitioners and researchers. This assessment tool could be implemented into music teacher education programs to help enhance building error detection skills among pre-service instrumental music educators. In addition, scholars could use the *IMEDT* to investigate discrepancies in what experiences contribute to error detection ability and what timbres and textures affect this ability.

Appendix A

University of Maryland Symphonic Wind Orchestra Instrumentation on IMEDT Recordings

Instrument	Number of Performers
Piccolo	1
Flute	4
Oboe	2
Clarinet	7
Bass Clarinet	1
Bassoon	2
Alto Saxophone	2
Tenor Saxophone	1
Baritone Saxophone	1
Horn	6
Trumpet	6
Trombone	3
Euphonium	2
Tuba	2
Percussion	5

Appendix B

Nature of Errors by Musical Excerpt

Musical Excerpt	Measure (Beat)	Type	Nature
A	1 (4)	R	Add extra eighth note in melody
A	9 (4)	A	Melody not slurring beats 4 and 5
B	82 (4)	P	Flutes play A natural instead of A flat
B	83 (3)	P	Flutes play A natural instead of A flat
B	84 (4)	R	Horns, Tenor Sax play last eighth note on 4 instead of on the and of 4
B	86 (3-4)	R	Trumpets play triplet instead of eighth note figure
B	88 (3)	P	Concert A natural instead of A flat (third of the chord) – makes chord major instead of minor
C	109 (2)	P	Trumpets below written E natural
C	111 (2)	R	Horns and Trombones enter on beat 1 instead of and of 2
C	115 (1-2)	A	Melody slurring instead of articulating eighth notes
C	115 (1)	P	Alto Sax playing F natural instead of F sharp
C	116 (1)	P	Alto Sax playing F natural instead of F sharp
D	18 (2)	P	Trumpet, Alto Sax, and Clarinet 1 play concert B flat instead of concert B natural
D	21 (1)	P	Clarinet and Alto Sax 2 play concert E flat instead of concert E natural

(Continued)

Musical Excerpt	Measure (Beat)	Type	Nature
D	22 (2)	A	Trombones articulate and of 2 instead of slur
D	24 (1-2)	R	Trumpet comes in on 1 instead of 2
E	30 (2)	R	Oboes, Clarinet 1, and Alto Sax 1 play on beat 2 instead of and of 2
E	33 (1-3)	A	Trombones articulate measure instead of slur
E	35 (2)	P	Trumpet, Clarinet, and Alto Sax 1 play concert B flat instead of concert B natural
E	36 (2-3)	R	Ensemble plays a 2/4 measure instead of 3/4
E	37 (3)	P	Clarinet 1 playing F natural instead of F sharp
F	35 (1-2)	A	Melody slurring instead of articulating
F	41 (2)	P	Melody playing concert A flat instead of A natural
F	42 – 43 (4-1)	R	Flute/Piccolo changing notes on beat 1 of 43 instead of beat 4 of 42
G	13 (6)	P	Bassoon and Bass Clarinet playing E flat instead of E natural
G	20 (1-2)	R	Melody playing dotted rhythm instead of straight eighths
G	21 (1-4)	R	Timpani plays only first beat
H	87 (1)	A	Trumpets articulate downbeat instead of tie over
H	88 (4-6)	R	Melody reverses the “quarter eighth” rhythm to “eighth quarter”
H	90 (1)	R	Baritone enters one measure early

Appendix C

Pitch Errors

Excerpt	Measure/Beat	Nature of Error
B	Meas. 82/4	Flutes play A natural instead of A flat
B	Meas. 83/3	Flutes play A natural instead of A flat
B	Meas. 88/3	Concert A natural instead of A flat (third of the chord) – makes chord major instead of minor
C	Meas. 109/2	Trumpets below written E natural
C	Meas. 115/ 1	Alto Saxophone playing F natural instead of F sharp
C	Meas. 116/1	Alto Saxophone playing F natural instead of F sharp
D	Meas. 18/2	Trumpet 1, Alto Saxophone 1, Clarinet 1 play concert B flat instead of concert B natural
D	Meas. 21/1	Clarinet 2 and Alto Saxophone 2 play concert E flat instead of E natural
E	Meas. 35/2	Trumpet 1, Clarinet 1, Alto Saxophone 1 play concert Bflat instead of concert B natural
E	Meas. 37/3	Clarinet 1 playing F natural instead of F sharp
F	Meas. 40/4	Melody plays a concert A flat instead of a concert A natural
G	Meas. 13/6	Bassoon and Bass Clarinet pay concert E flat instead of E natural

Appendix D

Rhythm Errors

Excerpt	Measure/Beat	Nature of Error
A	Meas.1/4	Extra eighth note added in melody
B	Meas. 84/4	Horns and Tenor Saxophone play last eighth note on 4 instead of the and of 4
B	Meas. 86/3-4	Trumpets hold tied note too long, therefore making eighth note figure into a triplet figure
C	Meas. 111/2	Horns & Trombones enter on beat 1 of measure 112 instead of on the and of 2 in measure 111
D	Meas. 24/2 (1)	Trumpets enter on beat 1 instead of beat 2
E	Meas. 30/2	Oboe, Clarinet 1, and Alto Saxophone 1 play on beat 2 instead of the and of 2
E	Meas. 36/2-3	Entire ensemble adds extra beat, creating a 3/4 measure instead of a 2/4 measure
F	Meas. 42-43/4-1	Piccolo and Flute play quarter note on beat 1 of measure 43 instead of beat 4 of measure 42
G	Meas. 20/2-3	Melody plays beat 2 as a dotted eighth, rather than straight eighths
G	Meas. 21/1	Timpani rhythm written as dotted eighth, but played as only quarter note on downbeat
H	Meas. 88/4-6	Melody reverses the “quarter eighth” rhythm to “eighth quarter”
H	Meas. 91/1	Baritone enters one measure early

Appendix E

Articulation Errors

Excerpt	Measure/Beat	Nature of Error
A	Meas. 9/4	Melody tonguing instead of slurring beats 4 and 5
C	Meas. 115/1 – 2	Melody slurring instead of articulating eighth notes
D	Meas. 22/2	Trombones articulate the and of 2 instead of slur
E	Meas. 33/1-3	Trombones articulate instead of slur entire measure
F	Meas. 35/1-2	Melody slurring beats 1 and 2 instead of tonguing
H	Meas. 87/1	Trumpets rearticulate beat 1 instead of holding tie from previous measure

Appendix F

Final Answer Sheet for IMEDT in Six Methods

Sample

Measure & Beat #	**What type of error? P=pitch, R=rhythm, A=articulation**	**What is the nature of the error? (please be as specific as possible)**
M35/B3	P	Bassoon playing Bb instead of B natural
M37/B1	R	Trumpets playing "1&2&" instead of "1&_&)
M38/B4	A	Upper woodwinds articulating eighth notes instead of slurring

Example 1

Measure & Beat #	**What type of error? P=pitch, R=rhythm, A=articulation**	**What is the nature of the error? (please be as specific as possible)**

Example 2

Measure & Beat #	**What type of error? P=pitch, R=rhythm, A=articulation**	**What is the nature of the error? (please be as specific as possible)**

Appendix G

Demographic Information Form

My name is Karen Koner and I am a doctoral candidate in the music education department at the University of Maryland. I am currently working on my dissertation under my advisor, Dr. Michael Hewitt. For my research, I have developed an error detection tool to examine a music educator's ability to detect errors in an instrumental ensemble setting.

I am hoping you could help me with my study. The procedure involves participants to complete an error detection tool that I recently developed. This process takes no more than 45 minutes. My assistant, Carolyn Sweterlitsh, or I will come to your school, or a place of your convenience, to administer the error detection tool in person during a time that works best for you.

As a thank you for your participation in this study, you will automatically be entered in a drawing for one of four gift cards to Amazon.com or iTunes

There are no known risks or benefits from participating in this research study.

Any potential loss of confidentiality will be minimized by using an online password-protected site for collecting data and in person testing. This information will be downloaded only to the researcher's computer, which is password protected.

In any presentation of this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.

The University of Maryland does not provide any medical, hospitalization or other insurance for participants in this research study, nor will the University of Maryland provide any medical treatment or compensation for any injury sustained as a result of participation in this research study, except as required by law.

Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator, Karen Koner or my advisor, Michael Hewitt at:

University of Maryland School of Music
2110 CSPAC (Room 2130D)
College Park, MD 20742
301-314-2490
kmkoner@umd.edu
mphewitt@umd.edu

If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:

University of Maryland College Park
Institutional Review Board Office
1204 Marie Mount
College Park, Maryland, 20742
E-mail: irb@umd.edu
Telephone: 301-405-0678

This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.

Clicking the box below indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. Please print this form for your records.

Clicking the box below indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you

voluntarily agree to participate in this research study. Please print this form for your records.

Name

Please state the number of years you have been an instrumental music educator in a K-12 setting (include full and part time years. Please include the current school year)

What is the primary grade level in which you currently teach instrumental music?

- Elementary (K - 5th Grade)
- Middle (6th - 8th Grade)
- High (9th - 12th Grade)
- I do not teach instrumental music in a K-12 setting
- I am still working on my undergraduate degree
- I do NOT teach instrumental music primarily in one grade level (example: 50% elementary and 50% middle school). Please Explain Below

In how many classes have you used instrumental score reading (e.g., conducting class, fieldwork experiences)?

- 0
- 1
- 2
- 3
- 4
- 5 or more

Please briefly describe these class experiences with instrumental score reading

In your current professional role, how often do you study instrumental music scores?

- Never
- A few times each year
- A few times each month
- A few times each week
- Once a day during the work week
- A few times a day during the work week
- Once a day, every day
- A few times a day, every day

What is your primary instrument (The instrument in which you primarily performed on during your undergraduate career)?

In what type of ensemble did you primarily perform as an undergraduate?

- Choir
- Band
- Orchestra
- Other (please specify)

How many hours a week do you practice/perform on your primary instrument outside of any K-12 teaching (excluding private lesson instruction)?

How many hours a week do you teach private lessons on your primary instrument outside of any K-12 teaching?

What is your highest earned degree?

- Bachelors
- Masters
- Doctorate
- Will graduate with Bachelor's at the end of the semester
- Other (please specify)

What is the content area of your highest earned degree (If you have not yet graduated with your bachelor's degree, what is the content area of the degree you will earn at the end of your current schooling)?

- Music
- Music Education
- Music Performance
- Conducting
- Music Theory
- Composition
- Musicology
- Ethnomusicology
- Elementary Education
- Secondary Education
- Educational Leadership
- Other (please specify)

What is the content area of your undergraduate degree (If you have not yet graduated with your bachelor's degree, what is the content area of the degree you will earn at the end of your current schooling)?

- Music
- Music Education
- Music Performance
- Conducting
- Music Theory
- Composition
- Musicology
- Ethnomusicology
- Elementary Education
- Secondary Education
- Educational Leadership
- Other (please specify)

What was your approximate GPA for your undergraduate degree upon graduation? (On a 4 point scale)

What was your overall aural skills/music theory grades earned during your undergraduate degree?

- A
- B
- C
- D or below

Appendix H

Item Analysis of Inserted Errors for Pilot Study

Examples A and B

Example & Measure of Error	**A1**	**A9**	**B82**	**B83**	**B84**	**B86**	**B88**
Participants in Pilot Study who Determined Error (*N = 17)*	9	1	15	13	4	12	8
Participants in Pilot Study who Determined Nature of Error (*N = 17)*	9	1	12	13	3	10	7

Examples C and D

Example & Measure of Error	**C109**	**C111**	**C115 (A)**	**C115 (P)**	**C116**	**D18**	**D21**	**D22**	**D24**
Participants in Pilot Study who Determined Error (*N = 17)*	10	2	3	13	3	15	10	3	4
Participants in Pilot Study who Determined Nature of Error (*N = 17)*	11	1	3	11	2	13	10	3	4

Examples E and F

Example & Measure of Error	**E30**	**E33**	**E35**	**E36**	**E37**	**F35**	**F41**	**F42**
Participants in Pilot Study who Determined Error (*N = 17)*	7	0	15	7	13	3	11	6
Participants in Pilot Study who Determined Nature of Error (*N = 17)*	7	0	13	6	9	2	5	6

Examples G and H

Example & Measure of Error	G13	G20	G21	H87	H88	H90
Participants in Pilot Study who Determined Error (*N = 17)*	13	10	5	4	9	2
Participants in Pilot Study who Determined Nature of Error (*N = 17)*	7	9	5	4	9	2

Appendix I

General Directions for All Six Test Variations

> "Thanks so much for participating today. For my dissertation for my PhD at Maryland, I decided to examine what contributes to the building of a music educator's error detection skills. As I was developing this research project, I examined several different error detection tests and decided that none of them accurately portrayed what happens in an instrumental music classroom, with examples being only one, two, or three parts, or recordings made from MIDI files. Therefore, I decided to create my own error detection test that uses a real ensemble making real mistakes to have a test that is as close to a "real life" teaching situation as possible.
>
> I have developed six different methods of taking this test to determine which one is the most valid and reliable to use when examining error detection skills. Today, you will use all six different methods when examining eight different musical excerpts. You will be able to examine the score prior to listening, you will be able to examine the score while listening, and you will just hear a model performance before examining the scores. With each of three methods, you will timed, or you will be allowed as much time as needed. As each excerpt approaches, I will direct you as to how to proceed through that individual excerpt."

"Did you fill out the questionnaire through SurveyMonkey (surveymonkey.com)?"

If not, hand participant survey

Hand Participant Answer Sheet:

"A few notes about the test in general, which are also stated on the front of you answer sheet":

- Each example will contain anywhere from 2 – 5 errors in pitch, rhythm, or articulation.
    - Pitch Errors = A note being performed different than pitch that is written
    - Rhythm Errors = manipulating the original rhythms in one of two ways. One way was the pitch was either held longer than indicted in the musical score or shorter than was indicated. The second was by performers playing before the rhythm was indicated to be played (early) or after the rhythm was indicated to be played (late).
    - Articulation errors = slurs and tonguing

- If the error is more than one beat long, please list the beat the note starts on (example: half note on beat 2 was held for 1 count too long, 3 counts. Therefore the error occurred on beat 2)
- In each example, the instruments Eb clarinet, alto clarinet, contrabass clarinet, and string bass have been eliminated
- If an example is in 6/8 – please indicated the beat number as the micro beat (1,2,3,4,5,6)
- Please do not include any errors if you do not hear them
- You will hear each altered performance twice.

Also remember:

- Please feel free to write on the score
- Fill in as much as you can on answer sheet
- Remember that at any time if you don't want to participate, you are free to withdraw from the study
- The only person that will be examining your answers is the researcher, Karen Koner

Individual Test Procedures

Test 1

Example 1(C): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 2 (D): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed

3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 3 (E): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 4 (F): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there not will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 5 (G): You will not hear a model performance and have one minute to examine the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Hand participant score
2. Look over score for up to 1 minute
3. Play altered recording
4. Have 1 minute to make notes
5. Play altered recording a second time
6. Have 2 minutes to complete answer box
7. Have participant turn score over after 2 minutes

Example 6 (H): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will be no time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long a s needed
5. Play altered recording a second time
6. Complete answer box

Example 7 (A): You will not hear a model performance and have one minute to examine the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Hand participant score
2. Look over score for up to 1 minute
3. Play altered recording
4. Have 1 minute to make notes
5. Play altered recording a second time
6. Have 2 minutes to complete answer box
7. Have participant turn score over after 2 minutes

Example 8 (B): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time
6. Complete answer box

Test 2

Example 1 (B): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 2 (C): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 3 (D): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 4 (E): You will not hear a model performance and have one minute to examine the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Hand participant score
2. Look over score for up to 1 minute
3. Play altered recording
4. Have 1 minute to make notes
5. Play altered recording a second time
6. Have 2 minutes to complete answer box
7. Have participant turn score over after 2 minutes

Example 5 (F): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time

6. Complete answer box

Example 6 (G): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 7 (H): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 8 (A): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Test 3

Example 1 (E): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 2 (F): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 3 (G): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 4 (H): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version

4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 5 (A): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 6 (B): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 7 (C): You will not hear a model performance and have one minute to examine the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Hand participant score
2. Look over score for up to 1 minute
3. Play altered recording
4. Have 1 minute to make notes
5. Play altered recording a second time
6. Have 2 minutes to complete answer box
7. Have participant turn score over after 2 minutes

Example 8 (D): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score

2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time
6. Complete answer box

Test 4

Example 1 (D): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 2 (E): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 3 (F): You will not hear a model performance and have one minute to examine the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Hand participant score
2. Look over score for up to 1 minute
3. Play altered recording
4. Have 1 minute to make notes
5. Play altered recording a second time
6. Have 2 minutes to complete answer box
7. Have participant turn score over after 2 minutes

Example 4 (G): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time
6. Complete answer box

Example 5 (H): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 6 (A): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time
6. Complete answer box

Example 7 (B): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance

5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 8 (C): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Test 5

Example 1 (A): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 2 (B): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 3 (C): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 4 (D): You will not hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to altered performance
4. Have 1 minute to make notes
5. Listen to altered performance a second time
6. Have 2 minutes to complete answer box
7. Turn score over after 2 minutes

Example 5 (E): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long a s needed
5. Play altered recording a second time
6. Complete answer box

Example 6 (F): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes

6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 7 (G): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Example 8 (H): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Test 6

Example 1 (F): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Have 1 minute to make notes on score
5. Listen to altered version again
6. Have 2 minutes to complete answer box

Example 2 (G): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 3 (H): You will not hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to altered performance
4. Have 1 minute to make notes
5. Listen to altered performance a second time
6. Have 2 minutes to complete answer box
7. Turn score over after 2 minutes

Example 4 (A): You will hear a model performance before examining the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Listen to model performance
2. Hand participant the score
3. Listen to altered version
4. Take as much time as needed to make notes
5. Listen to altered version again
6. Complete answer box

Example 5 (B): You will not hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to altered performance
4. Have 1 minute to make notes
5. Listen to altered performance a second time
6. Have 2 minutes to complete answer box
7. Turn score over after 2 minutes

Example 6 (C): You will not hear a model performance and have as much time as needed to examine the score. This will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Hand participant score
2. Look over score for as long as needed
3. Play altered recording
4. Take notes for as long as needed
5. Play altered recording a second time
6. Complete answer box

Example 7 (D): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will be time constraints.

1. Give participant score
2. Look over score for up to 1 minute
3. Listen to model performance
4. Listen to altered performance
5. Have 1 minute to make notes
6. Listen to altered performance a second time
7. Have 2 minutes to complete answer box
8. Turn score over after 2 minutes

Example 8 (E): You will hear a model performance while examining the score, which will be followed by hearing the altered performance twice. On this example there will not be time constraints.

1. Give participant score
2. Look over score for as much time as needed
3. Listen to model performance
4. Listen to altered performance
5. Make notes for as much time as needed
6. Listen to altered performance a second time
7. Complete answer box

Appendix J

Initial Email Invite to Potential Participants

Hi (Insert name here),

I hope you are having a good semester and looking forward to the holiday break! As I am now a doctoral candidate in the music education department at the University of Maryland, I am currently working on my dissertation under my advisor, Dr. Michael Hewitt. For my research, I have developed an error detection tool to examine a music educator's ability to detect errors in an instrumental ensemble setting.

I am hoping you could help me with my study. The procedure involves participants to complete an error detection tool that I recently developed. This process takes no more than 45 minutes. My assistant, Carolyn Sweterlitsch, or I will come to your school, or a place of your convenience, to administer the error detection tool in person during a time that works best for you. I am hoping to collect all data by the end of January, 2012.

There are no known risks from participating in this research study. However, as thanks for your participation in this study, you will automatically be entered in a drawing for one of five gift cards to Amazon.com or iTunes worth $25 each.

I do hope that you will consider taking the time to be a part of this study that will help inform the development and refinement of error detection skills amongst musicians and music educators. Please let me know as soon as possible if you would be able to participate. You can also reach me on my cell phone at [redacted]

Thanks so much for your time and I hope to hear from you soon,

Karen Koner

Appendix K

Answer Sheet for IMEDT in S&R/N Method

Error Detection

For the following listening excerpts, please fill in the chart to the best of your ability.

- Each example will contain anywhere from 2 – 5 errors in pitch, rhythm, or articulation.
    - Pitch Errors = A note being performed different than pitch that is written
    - Rhythm Errors = manipulating the original rhythms in one of two ways. One way was the pitch was either held longer than indicted in the musical score or shorter than was indicated. The second was by performers playing before the rhythm was indicated to be played (early) or after the rhythm was indicated to be played (late).
    - Articulation errors = slurs and articulations
- Fill in as much as you can
- In each example, the instruments Eb clarinet, alto clarinet, contrabass clarinet, and string bass have been eliminated
- If the error is more than one beat long, please list the beat the note starts on (example: half note on beat 2 was held for 1 count too long, 3 counts. Therefore the error occurred on beat 2)
- If an example is in 6/8 – please indicated the beat number as the micro beat (1,2,3,4,5,6)
- The approximate tempo for each example is indicated at the top of the answer key
- Please do not include any errors if you do not hear them
- You will hear each altered performance twice.
- Please feel free to write on the score

Sample

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)
M35/B3	P	Bassoon playing Bb instead of B natural
M37/B1	R	Trumpets playing “1&2&” instead of “1&_&)
M38/B4	A	Upper woodwinds articulating eighth notes instead of slurring

Example 1 (E) – Largo (quarter note = 56)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 2 (F) – Allegro (quarter note = 138)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 3 (G) – Andante (dotted quarter note = 92)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 4 (H) – Allegro (dotted quarter note= 120)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 5 (A) – Allegro (dotted quarter note = 116)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 6 (B) – Largo (quarter note = 66)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 7 (C) - Allegro (dotted quarter note = 116)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

Example 8 (D) - Largo (quarter note = 56)

Measure & Beat #	What type of error? P=pitch, R=rhythm, A=articulation	What is the nature of the error? (please be as specific as possible)

References

Battisti, F. L. (2002). *The winds of change: The evolution of the contemporary American wind band/ensemble and its conductor*. Galesville, MD: Meredith Music.

Battisti, F. & Garofalo, R. (1990). *Guide to score study for the wind band conductor.* Galesville, MD: Meredith Music.

Blocher, L. R. (1986). *An analysis of college band instrumentalists' detection of common performance errors* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 8708160).

Boyle, J. D. & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences.* New York: Schirmer Books.

Brand, M. & Burnsed, V. (1981). Music abilities and experiences as predictors of error-detection skill. *Journal of Research in Music Education, 29,* 91-96. doi: 10.2307/3345017.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10*(2), 161 – 180. doi: 10.1207/s15324818ame1002_4.

Byo, J .L. (1993). The influence of textural and timbral factors on the ability of music majors to detect performance errors. *Journal of Research in Music Education, 41*, 156-167. doi: 10.2307/3345405.

Byo, J. L. (1997). The effects of texture and number of parts on the ability of music majors to detect performance errors. *Journal of Research in Music Education, 45*, 51-66. doi: 10.2307/3345465.

Byo, J. L. & Sheldon, D. A. (2000). The effect of singing while listening on undergraduate music majors' ability to detect pitch and rhythm errors. *Journal of Band Research, 36*(1), 26-46.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminat validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105. doi: 10.1037/h0046016.

Colwell, R. J. (1969) *Music achievement tests 1 and 2.* Chicago: Follett Educational Corporation.

Colwell, R. J. & Hewitt, M. P. (2011). *The teaching of instrumental music* (4th ed.). Upper Saddle River, NJ: Pearson.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Upper Saddle River, NJ: Pearson.

Crowe, D. R. (1996). Effects of score study on beginning conductors' error detection abilities. *Journal of Research in Music Education, 44*, 160-171. doi: 10.2307/3345668.

DeCarbo, N. J. (1981). *The effects of conducting experience and programmed materials on error detection scores of college conducting students*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 8121086).

DeCarbo, N. J. (1982). The effects of conducting experience and programmed materials and error-detection scores of college conducting students. *Journal of Research in Music Education, 30*, 187-200. doi: 10.2307/3345085.

Doane, C. (1988/1989). The development and testing of a program for the development of conductor aural error detection skills. *College Band Directors National Association Journal, 6*, 11-14.

Doerksen, P. F. (1999). Aural-diagnostic and prescriptive skills of preservice and expert instrumental music teachers. *Journal of Research of Music Education, 47*, 78-88. doi: 10.2307/3345830.

Education Testing Services (2003). *Study guide for the music tests: Concepts and process; Analysis; and content knowledge.* Washington DC: Educational Testing Services.

Educational Testing Services. (n.d.). Retrieved from http://www.ets.org/praxis.

Farnum, S. E. (1969). *Farnum string scale.* Winona, MN: Hal Leonard.

Faul, F., Erdfelder, E. Lang, A., & Buchner, A. (2007). G Power (Version 3.1) [Computer software].Germany. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education, 25*, 256-263. doi: 10.2307/3345266

Forsythe, J. L. & Woods, J. R. (1983). The effects of conducting on the error detection ability of undergraduate and graduate instrumental conductors. *Contributions to Music Education, 10*, 27-32.

Froseth, J. 0., & Grunow, R. A. (1979). *MLR instrumental score reading program*. Chicago: G.I.A. Publications.

Garofalo, R. J. & Whaley, G. (1979). Comparison of the unit study and traditional approaches for teaching music through school band performance. *Journal of Research in Music Education, 27*, 137-142. doi: 10.2307/3345266.

Goolsby, T. W. (1996). Time use in instrumental rehearsals: A comparison of experienced, novice, and student teachers. *Journal of Research in Music Education, 44*, 286-303. doi: 10.2307/3345442.

Goolsby, T. W. (1997). Verbal instruction in instrumental rehearsals: A comparison of three career levels and preservice teachers. *Journal of Research in Music Education, 45*, 21-40. doi: 10.2307/3345463.

Goolsby, T. W. (1999). A comparison of expert and novice music teachers' preparing identical band compositions: An operational replication. *Journal of Research in Music Education, 47*, 174-187. doi: 10.2307/3345722.

Gonzo, C. L. (1971). An analysis of factors related to choral teachers' ability to detect pitch errors while reading the score. *Journal of Research in Music Education, 19*, 259-271. doi: 10.2307/3343762.

Gordon, E. (1965). *Musical aptitude profile.* Boston: Houghton Mifflin.

Gordon, E. (1970). *Iowa tests of music literacy.* Iowa City: Bureau of Educational Research and Service.

Gordon, E. (1978). *Primary measures of music audiation.* Chicago: G.I.A. Publications.

Gordon, E. (1982). *Intermediate measures of music audiation.* Chicago: G.I.A. Publications.

Gordon, E. (1989). *Advanced measures of music audiation.* Chicago: GIA Publications.

Grunow, R. L. (1980). *An investigation of the relative effectiveness of four modes of score preparation on visual-aural discrimination skills development.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 8017275).

Harrison, C. R., 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*, 131-144. doi: 10.2307/3345497.

Hewitt, M. P. (2007). Influence of primary performance instrument and education level on music performance evaluation. *Journal of Research in Music Education, 55*, 18-30. doi: 10.1177/002242940705500103.

Hewitt, M. P. & Smith, B. P. (2004). The influence of teaching-career level and primary performance instrument on the assessment of music performance. *Journal of Research in Music Education, 52*, 314-327. doi: 10.1177/002242940405200404.

Hochkeppel, W. (1993). *Systematic score study: Effects of four methodologies on error detection achievement in instrumental conducting students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 9402969).

Hopkins, J. (1991). *The effect of four approaches to score study on student conductors' ability to detect errors in the performance of choral music.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 9210841).

Humphreys, J. T. (1993). Precursors of musical aptitude testing: From the Greeks through the work of Francis Galton. *Journal of Research in Music Education, 41*, 315-327. doi: 10.2307/3345500.

Jennings, D. L. (1988). *The effectiveness of instrumental music teacher pre-service training experiences as perceived by college and high school band directors.* (Doctoral dissertation). Available from ProQuest (AAT 8910137).

Killian, J. N. (1991). The relationship between sightsinging accuracy and error detection in junior high singers. *Journal of Research in Music Education, 39*(3), 216-224. doi: 10.2307/3344721.

Kostka, S. & Payne, D. (2009). *Tonal Harmony* (6th ed.). New York, NY: McGraw-Hill Higher Education.

La Plante, P. (1991). *American riversongs: A folksong setting for band* [Score]. New Glarius, WI: Daehn Publications.

Larson, R. C. (1977). Relationships between melodic error detections, melodic dictation, and melodic sightsining. *Journal of Research in Music Education, 25*(4), 264-271. doi: 10.2307/3345267.

Locy, R. S. (1996). *The effect of instrumental timbre preference and instrumental timbre on the pitch error detection skills of university conducting students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 9707232).

Lomax, R. G. (2007). *Statistical concepts* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malone, P. (1985). *Development and evaluation of an approach to increasing pitch error detection skills in choral music education students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 8602871).

Margolis, B. (1993). *The Renaissance fair* [Score]. Brooklyn, NY: Manhattan Beach Music.

Maslanka, D. (1985). *Rollo takes a walk* [Score]. Park Ridge, IL: Neil A Kjos Music Company.

Messick, S. (1989). Validity. In R. L. Linn (Editor), *Educational Measurement* (3rd Ed.). New York, NY: Macmillan Publishing Company.

Miles, R. (Ed.). (1997). *Teaching Music Through Performance in Band* (Vol. 1). Chicago: GIA Publications.

Miles, R. (Ed.). (2000). Teaching Music Through Performance in Band (Vol. 3). Chicago: GIA Publications.

Miles, R. (Ed.). (2004). Teaching Music Through Performance in Band (Vol. 5). Chicago: GIA Publications.

National Association of Schools of Music (2012). In *National Association of Schools of Music Handbook* (2011-2012 ed.). Retrieved from http://nasm.arts-accredit.org.

Ramsey, D. S. (1978). *Programmed instruction using full-score band literature to teach pitch and rhythm error detection skill to college music education students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 7902932).

Ramsey, D. S. (1979). Programmed instruction using band literature to teach pitch and rhythm error detection to music education students. *Journal of Research in Music Education, 27*, 149-162. doi: 10.2307/3344966.

Seashore, C. E. (1915). The measurement of musical talent. *The Musical Quarterly, 1*, 129-148. Retrieved from http://www.jstor.org/stable/10.2307/738047.

Sheldon, D. A. (1998). Effects of contextual sight-singing and aural skills training on error-detection abilities. *Journal of Research in Music Education, 46*, 384-395. doi: 10.2307/3345550.

Sheldon, D. A. (2004). Effects of multiple listenings on error-detection acuity in multivoice, multitimbral musical examples. *Journal of Research in Music Education, 52*, 102-115. doi: 10.2307/3345433.

Sibelius (Version 3) [Computer software]. (2003). Available from http://www.sibelius.com/home/index_flash.html.

Sidnell, R. G. (1971). Self-Instructional drill materials for student conductors. *Journal of Research in Music Education, 19*, 85-91. doi: 10.2307/3344117.

Snow, R. E. (1974). Representative and quasi-representative designs for research on teaching. *Review of Educational Research, 44*(3), 265-291. Retrieved from http://www.jstor.org/stable/1169976.

Straub, D. W. Validating instruments in MIS research. *MIS Quarterly, 13*, 147-169.

Stuart, M. (1979). The use of videotape recordings to increase teacher trainees' error detection skills. *Journal of Research in Music Education, 27*, 14-19. doi: 10.1177/0022429410396094.

SurveyMonkey (1999-2012). [Computer software]. Available from http://www.surveymonkey.com/

Taebel, D. A. (1980). Public school music teachers' perceptions of the effects of certain competencies on pupil learning. *Journal of Research in Music Education,* 28, 185-197. doi: 10.2307/3345236.

Thornton, L. (1998). *The effects of listening condition on melodic error detection by novice woodwind students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (AAT 9904870).

Tichelli, F. (1991). *Cajun folk songs* [Score]. Brooklyn, NY: Manhattan Beach Music.

University of Maryland College of Education. (n.d.). Retrieved from http://www.education.umd.edu/EDCI/info/internship.html).

Urbaniak, G. C. & Plous, S. (1997-2012). Research randomizer (Version 3.0) [Computer software]. Available from October, 2011, from http://www.randomizer.org/

Van Oyen L. G. & Nierman, G. (1998). The effects of two approaches to instrumental score preparation on the error detection ability of student conductors. *Contributions to Music Education*, *25*(1), 85-97.

Waggoner, D. T. (2011). Effects of listening conditions, error types, and ensemble textures on error detection skills. *Journal of Research in Music Education, 59*, 56-71.

Watkins, J. G. & Farnum, S. E. (1954). *The Watkins-Farnum performance scale: Form a.* Winona, MN: Hal Leonard.

Watkins, J. G. & Farnum, S. E. (1962). *The Watkins-Farnum performance scale: Form b*. Winona, MN: Hal Leonard.

Williams, M. (1995). *Highbridge excursions* [Score]. Van Nuys, CA: Alfred Publishing Co.