



8-2000

## Assessment and Grading Practices Among High School Band Teachers in the United States: A Descriptive Study

Edward G. Simanton

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: <https://commons.und.edu/theses>



Part of the [Psychology Commons](#)

---

### Recommended Citation

Simanton, Edward G., "Assessment and Grading Practices Among High School Band Teachers in the United States: A Descriptive Study" (2000). *Theses and Dissertations*. 961.  
<https://commons.und.edu/theses/961>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact [und.common@library.und.edu](mailto:und.common@library.und.edu).

3 3100 02047 0265



UND CHESTER FRITZ LIBRARY

ASSESSMENT AND GRADING PRACTICES AMONG HIGH SCHOOL BAND  
TEACHERS IN THE UNITED STATES: A DESCRIPTIVE STUDY

by

EDWARD G. SIMANTON

Bachelor of Education, Union College, 1985  
Master of Music Education, Brandon University, 1996

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

August

2000

T2000  
55883

This dissertation, submitted by Edward G. Simanton, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

John A. Hoover  
(Chairperson)

Richard L. Landry

Katherine Norman

Mary Lou Fuller

Tom Petros

This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Harvey Kuhl  
Dean of the Graduate School  
7-27-00

\_\_\_\_\_  
Date


PERMISSION

Title            Assessment and Grading Practices Among High School Band Teachers in  
                    the United States: A Descriptive Study

Department    Teaching & Learning

Degree         Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation

Signature 

Date            7-18-2000

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
ACKNOWLEDGMENTS .....	x
ABSTRACT .....	xi
CHAPTER	
I. INTRODUCTION .....	1
History of Accountability in American Education .....	4
History of Instrumental Music Education .....	5
Rationale for the Study .....	8
The Problem .....	9
Purpose of the Study .....	9
Delimitations of the Study .....	10
Definition of Terms .....	10
Summary .....	11
II. REVIEW OF LITERATURE .....	12
Introduction .....	12
Musical Learning Assessment Methods .....	12
Actual Practices and Attitudes Toward These Practices .....	26
Definition of Assessment in Music Education .....	31

	Roles of Assessment .....	32
	Summary .....	33
III.	METHOD .....	34
	Introduction .....	34
	Selection of Participants .....	34
	Development of Surveys .....	37
	Procedures .....	37
	Analysis of the Data .....	38
	Summary .....	40
IV.	RESULTS .....	41
	Survey Response Data .....	41
	Demographic Information .....	43
	Summary of Assessment Data (Research Question One) .....	45
	Summary of Grading Data (Research Question Two) .....	48
	Summary of Assessment and Grading Satisfaction Data (Research Question Three) .....	51
	Relationships Between Demographic and Dependent Variables (Research Question Four) .....	53
V.	CONCLUSIONS .....	60
	Introduction .....	60
	Summary of Findings Regarding Research Questions .....	60
	Summary of Findings Regarding Best Practice .....	63
	Comparison with Past Studies .....	65

Implications .....	65
Recommendations for Further Research .....	68
APPENDICES .....	70
REFERENCES .....	79

## LIST OF TABLES

### Table

1. MENC Regions, States Included in Regions, Percentage of US High Schools by Region, and Number of Selected Schools by MENC Region .....	36
2. Survey Return Information by Region .....	42
3. Frequency of Individual Student Performance Assessment .....	46
4. Percentage of Band Directors Who Audio Tape and/or Video Record Student Performance .....	47
5. Percentage of Grading Criterion used in Student Grading .....	49
6. Percentage of Grading Criterion Weighting .....	50
7. Percentage of Band Directors Who Inform Students, Parents and Principals Regarding Band Grading Systems .....	51
8. Comparison of Current and "Should Be" Grading Criterion Weighting .....	52
9. Mean Response to Assessment and Grading Opinion Items in Descending Mean Order .....	54



## LIST OF FIGURES

### Figure

1. Distribution of student assessment duration categories in seconds ..... 46
2. Non-instrumental performance assessment use ..... 48
3. Reported grade distribution for band students ..... 51
4. Percent use of tape recorder based on band size and band  
director education ..... 56
5. Mean assessment duration by MENC region ..... 58
6. Comparisons between the McCoy (1991) study and the  
present study on the basis of grade weightings ..... 66

## ACKNOWLEDGMENTS

I would like to acknowledge those who contributed to this study and express my appreciation for their part in its development.

Dr. John Hoover, chairman of the advisory committee, for his unfaltering support, guidance, and enthusiasm.

Dr. Katherine Norman, for her insight and expertise in music education.

Dr. Richard Landry , for his wisdom and guidance in all phases of the study.

Dr. Mary Lou Fuller and Dr. Tom Petros for their support and interest as committee members.

Dr. Marilyn Klug for her gracious help and technical support.

I would like to thank the many band directors who took the time to complete and return surveys. Without their help, this study would have been impossible.

Finally I would like to thank my family: my daughter Alana, who was with me in most phases of the project (sleeping on my lap during the pilot study interviews); my son Grant, who helped me after school; and most of all my wife Sandy, who has believed in and supported me always.

## ABSTRACT

The recent movement in academic reform and accountability has brought assessment and grading to the forefront in most academic disciplines. While assessment and grading appear to be important topics in music education professional journals, little is known about actual practice or local satisfaction with current practices. Findings from a few small regional assessment and grading studies indicated that assessment practices tended to diverge from currently understood best practices; for example, attendance was the primary source of grading information.

The purposes of this study were to (a) examine current assessment and grading practices in American high school bands, (b) gauge local satisfaction with current assessment and grading practices, and (c) investigate variations in practices and satisfaction based on regional, school, and band director variables.

Data were collected (via surveys) from 202 high school band directors using a regionally stratified sample, the six regions comprising the Music Educator's National Conference [MENC]. Findings indicate that while few band directors' assessment resembles what MENC representatives list as best practice and grades are made up primarily of non-musical criteria, subjects expressed a high degree of satisfaction with current practice.

Assessment was found to be closer to best practice in smaller bands and among band directors with graduate degrees. Time spent on assessment and use of grading

• criteria were found to vary regionally.

Further research was recommended to examine (a) the roles of assessment and grading in high school bands from the perspectives of students, parents, and principals, (b) the effectiveness of formal and informal assessment strategies, (c) the effect of band size and teacher background on assessment and grading, (d) factors influencing regional differences in assessment and grading practices, and (e) changes of assessment and grading over time.

## CHAPTER I

### INTRODUCTION

As more and more emphasis is placed on student performance and teacher accountability, measurement and assessment are becoming increasingly important to all music educators. With the inclusion of music as a core subject in the Goals 2000: Educate America Act, it is critical that music educators possess not only comprehensive knowledge of the subject matter but also the ability to assess the learning of that subject matter.

(Cope, 1996, p. 39)

The recent movement for academic reform and accountability is unlike any other education movement in American history. Although efforts to improve schooling can be traced to the founding years of this nation, the debates for educational accountability have reached national proportions in the post-Sputnik years (Fullan, 1993).

Accountability was addressed in the National Governors Conference in 1989 that laid the groundwork for the Goals 2000: Educate America Act. This formal declaration, which Bill Clinton signed into law on March 31, 1994, cited English, mathematics, history, geography, foreign language, science, civics and government, and the arts as "basic" to an American education. According to the document, students at grades 4, 8 and 12 are to demonstrate competence in each of these subject areas by the year 2000. Although the bill calls for the development of national standards in each discipline, the

more difficult challenge is the development of educational components necessary to ensure their success (American Council for the Arts, 1995).

The developers of the National Standards for Arts Education insist that an education in the arts be a comprehensive, sequenced enterprise in learning (Wolverton, 1992). Besides the performance techniques, which are often the primary educational objective of today's music programs, students should demonstrate knowledge and appropriate upper-level cognitive skills. For example, students should understand the musical forms and harmonic techniques being performed and evaluate their role within the historical context of the musical works. The completion of America's first national K-12 arts education standards was announced by the National Committee on Standards in the Arts in January of 1994 (MENC, 1994).

One of the major obstacles that must be faced in order to successfully meet the challenge of the national standards in music is in the area of assessment. According to Shepard (1989), assessment should be designed to resemble authentic learning tasks and support instruction. Assessment should also be sensitive enough to detect short-term changes and be relevant to local situations. Assessment should also be scored locally and provide meaningful feedback.

Crooks (1988) concluded from a review of educational research that the following are six ways in which assessment influences education: (a) Student response to assessment affects teacher pedagogy, (b) assessment guides students' judgment of what is important, (c) it affects student motivation and self-perception of competence, (d) it structures personal study time, (e) it consolidates learning, (f) it affects lifetime learning strategies. Although many music educators are pleased with the important place given to

music within educational reform movements, there is concern that music educators may not be prepared for the role they must play.

Music educators have developed many formative and summative assessment methods that can be applied to large ensemble rehearsals including standardized tests (Zdzinski, 1996), use of audio/videotape (Carlin, 1996; Killian, 1998; MENC, 1996) and rating scales (Cope, 1996; Robinson, 1995; Swanwick, 1994). Although music educators have applauded these developments, they are seldom used in practice (Lehman, 1992).

Despite the development of appropriate assessment tools and pressure to reform measurement and assessment in music education, high school music teachers' grading still tends to be based primarily on attendance (Lehman, 1992; McClung, 1997). This indifference to accepted wisdom is the result of many factors.

First, music education is not supervised with the same rigor as other subjects such as math, reading, and science. In fact, "no arm of the federal government exerts any jurisdiction over it and very few states do anything more than make helpful suggestions" (Britton, 1991, p. 178).

Second, music is considered to be more complicated and time consuming to measure than other subject areas. Although assessment methods have been developed, many music educators consider music performance assessment too time consuming or unrelated to program objectives for use in grading. Even with the use of rating scales, and other assessment tools, many music educators fear that elements such as tempo, phrasing nuances, use of ornamentation, and tone quality are creative expressions not appropriate for use in grading (Boyle & Radocy, 1987).

Even if music educators felt comfortable with more performance based grading, the student-to-teacher ratio in many high school music programs makes it impossible for teachers to find time to listen to students individually in addition to their excessive workloads. As stated by Robinson (1995), "Ensemble directors often find themselves dealing with large numbers of students, making the personal contact needed for in-depth learning to take place extremely difficult if not impossible" (p. 30).

Third, many music educators believe that group evaluation is an integral part of teaching performance skills and that these skills are already being demonstrated daily (Colwell, 1991). The individual evaluation is seen as redundant since individual performance is part of the ensemble performance.

To see where the field of music is, it is important to trace where it has been. Accordingly the history of assessment and education accountability is outlined in the next paragraphs.

#### History of Accountability in American Education

Thomas Jefferson is credited with the idea of providing educational opportunity to all citizens because he believed that no democratic society is safe without an educated population (Pulliam, 1991). After independence, education was viewed as a way for immigrants from many nations to become real Americans.

The era between the civil war and the first world war saw the development of modern educational systems. During this time period, schools became far less rigid, largely due to the work of individuals such as John Dewey. Courses in physical education, art, and music began to be offered in the late 19th century although, in many cases, credit was not given for electives until the 1920s (Montgomery, 1994).



As quality and class offerings increased, so did the cost of education. Even with inflation taken into account, spending per pupil increased 500% between 1945 and 1985 (Boyd & Kerchner, 1988). By shifting the financial burden of education, this increasing cost was not immediately noticed. Local support for education, which in 1930 was 83%, dropped to 51% in 1985 and 45% in 1996 (National Education Association, 1996). As the funding for education comes from more sources, so do its critics (Sarason, 1995).

Accountability movements in education have come at times when the security of the nation is thought to be at risk because of some perceived inability to compete globally. The Soviet launch of the first space capsule, Sputnik, led to concerns about American education. In answer to these concerns, American public education underwent a dramatic series of reforms aimed at improving student performance in math, science and foreign languages (Montgomery, 1994). Among the many similarities between Sputnik era reforms and the Goals 2000 program is the expectation for educational accountability (Flynn, 1995). Popham (1973) noted that not only does educational accountability require that measurable learning takes place but requires that the educator produce objective evidence that learning was the result of instruction (p. 107). The movement toward establishing that courses of study produce outcomes has also affected music education.

#### History of Instrumental Music Education

Instrumental music education in American public schools is a relatively new phenomenon. The first recorded use of instruments in schools was at the Boston Farm and Trades School in 1857. Instrumental music was not included in school curricula earlier because the early colonialists considered the social diversion of instrumental

music to be "frivolous if not wicked" (Birge, 1928, p. 173). This skeptical view of instrumental music education has since changed and "it is not an exaggeration to characterize school bands and orchestras, along with the proverbial motherhood and apple pie, as symbols of much that is good and wholesome in American life" (Humphreys, 1989, p. 50).

Instrumental music education, though present in a few scattered schools since the mid 1800s, was not widespread until the beginning of the 20th century. Early instrumental programs tended to be centered around orchestras rather than bands. In fact a study carried out in the 1919-1920 academic year by McConathy, Gehrkins, and Birge (1921) revealed that 278 out of 359 cities had school orchestras while only 88 cities had school bands. Even today, the time from the turn of the century until the early 1920s is considered the heyday of American public school orchestras.

By the end of World War I, military bands had become a symbol of American victory and with the war's end military trained bandleaders became available to the schools as music teachers. By the end of the 1920's most high schools and many grade schools included band in the curriculum.

Coinciding with the expansion of school music during the early 1900s was the new progressive education movement, which sought to increase high school enrollment and expand school offerings including the use of electives (Birge, 1928). This system made it possible for the first time ensembles to meet during school hours and for students to receive credit for music instruction.

In 1907, the Music Supervisor's National Conference was formed, later to become the Music Educator's National Conference [MENC] in 1934. Although this organization

was originally formed to oversee vocal music, it became a strong promoter of instrumental music in the 1920s and 1930s. In the early years of instrumental music, ensembles varied greatly depending on the available musicians and whims of local instructors. Almost any heterogeneous group of wind instruments could be called a band until MENC formed the Committee on Instrumental Affairs (Schleuter, 1984), which recommended instrumentation standards. The Committee on Instrumental Affairs also assisted in dissemination of recommended teaching methods and organization of competitions.

National band contests began in the early 1920s with many difficulties. In 1926 the National School Band Association was formed to help administer future contests. Its name was changed in 1929 to the National School Band and Orchestra Association. In the early 1930s, national contests were seen as becoming too competitive and required transporting bands great distances. This led to a change to regional festivals, which although still competitive were seen as more educational. These festivals are still an important aspect of school band programs (Schleuter, 1984).

The competitive aspect of band festivals provides valuable evaluation and feedback to students and teachers. This may contribute to the perception that individual performance evaluation and grading is unnecessary and redundant. Festival scores are also used as informal measures of program quality although this was never the intention of music festival planners (Lehman, 1992). It is possible that many band directors' emphasis on ensemble, rather than individual, assessment is a result of band festivals which are based on ensemble rather than individual performance.

Although band programs have made remarkable progress in the last 150 years and are now accepted as a basic part of American education, they still lack the universal acceptance enjoyed by other subjects such as math, English, and science. Deficiencies in assessment and grading, either perceived or real, remain major obstacles to music attaining status as a core subject. For example, because of a tendency among teachers of large ensembles to grade on attendance, music grades are often disregarded by college admissions officials (Lehman, 1992).

#### Rationale for the Study

In the 1920s, a movement to reform evaluation in music education caused a lowering of standards and, for a time, reduced music education to classroom games such as music bingo, and name the rhythm tic-tac-toe (Colwell, 1991). Today, "American music education is better than we tend to think it is. . . . European children simply do not have readily available the wealth of musical instruction taken for granted here" (Britton, 1991, pp. 175, 179).

Much of what makes American music successful could be freedom from regulation. It is entirely possible that regulation, supervision, and assessment could have the same negative effect on current American music education that it did in the 1920s (Colwell, 1991). This does not mean that assessment is not a necessary component of modern music education. Rather it means that care must be taken to examine present assessment forms and functions before they are disregarded as unacceptable to the future needs of music educators.

At present, what little is known about assessment and grading in high school band programs is based on personal observation and four regional studies. If music educators

are to successfully make the decisions needed to ensure a continued place for music in the new accountability-heavy educational climate, accurate data regarding current assessment and grading practice of American high school bands must be made available. As these data are not available, I propose to gather them as part of this dissertation.

### The Problem

Although dissatisfaction exists with current high school band assessment and grading practices at a national level (Cope, 1996), that dissatisfaction is based only on limited regional data. More study is needed to accurately describe current band assessment and grading practices at a national level. In addition, satisfaction regarding band assessment and grading practices has not been examined. Before changes can be made at a national level, practices and perspectives must be clearly understood.

### Purpose of the Study

In this study, I examined current assessment and grading practices in United States high school bands and local satisfaction with those practices. The questions to be answered by this study were as follows:

1. What do high school band directors report doing (what strategies and how frequently) to assess student learning within their band programs?
2. What factors are reported to be used (and in what percentages) in high school band grade assignment?
3. To what degree are the high school band assessment and grading practices viewed as adequate and appropriate from the band director's perspective?
4. Which of the following factors may be associated with variations in reported assessment and grading practices and/or perspectives regarding those practices: regional

factors (MENC region, urban versus rural), school factors (school size, number of students per band), band director factors (educational background, years experience, major instrument)?

#### Delimitations of the Study

This study was conducted with a regionally stratified random sample of public high school band directors in the United States. Validity of findings is dependent upon the accuracy of data provided by volunteer respondents. Data provided by high school band director respondents may not necessarily represent the responses of all American high school band directors. Findings should not be generalized beyond the United States or to private schools or grade levels other than those addressed.

#### Definition of Terms

Assessment refers to the systematic gathering of information and judgment based on that information to appraise individual student achievement. Going one step further, in this study, the term assessment refers only to the gathering and judgment of information regarding individual performance not the ensemble as a whole. As such, the spontaneous listening and reacting process used in rehearsals was not considered to constitute assessment.

Grading refers to the process of reporting information to parents, students, and school officials. This included letter grades, verbal descriptions, numeric data, and portfolios.

Student refers to students enrolled in large instrumental ensembles (more than 25 players) at the high school level. These large ensembles can be distinguished from other ensembles both by their size and their general performance usage.

### Summary

Recent educational and political movements have focused attention on assessment and accountability in education, including music education. While many music educators applaud inclusion of music education as a core subject in the Goals 2000: Educate America Act as an indication that music education will play an important role in the future of American public education, many also fear that, at present, music educators may not be equipped to meet the assessment needs of these educational movements (Cope, 1996).

Since the early years of public school music education, teachers have struggled with assessment from the early days of the national band contests to the reform movement of the 1920s that caused a lowering of standards. While much has been written regarding assessment in the recent music education literature, little is known about actual practice. As music educators face the challenges of modern educational movements, important decisions must be made to ensure the future of public school music education. If music educators are to successfully make the decisions needed to ensure a continued place for music in the new accountability-heavy educational climate, accurate data regarding current assessment and grading practice must be made available.

## CHAPTER II

### REVIEW OF LITERATURE

#### Introduction

The purpose of this study was to examine assessment practices in high school band programs and the attitudes toward those practices from the perspective of band directors. In this chapter, I will review literature related to the present work focusing on studies that describe and evaluate musical learning assessment methods and in which music assessment practices or attitudes toward music assessment were investigated. A major section will be devoted to each. The chapter will conclude with overall findings as they relate to the current study.

#### Musical Learning Assessment Methods

Because of the many differences between music and traditional subjects, such as math and English, many different music assessment methods have been developed. The following is an examination of published descriptions of these music assessment methods, standardized tests, use of technology, and performance based music assessment based on measurement rubrics. Because these methods lend reliability and validity to music assessment their frequency of use and mode of application directly address the research questions introduced in Chapter One.



### Standardized Music Tests

Many standardized tests are available to secondary music educators. Although most of the well known standardized music tests were developed in the late 1960s and early 1970s, newer editions of many tests are available. The Iowa Tests of Music Literacy was first published in 1970 but was revised in 1991 (Gordon, 1970, 1991). This test purports to measure music aptitude, by testing tonal and rhythmic audiation, and music achievement, by testing music reading and vocabulary, in fourth through twelfth grade students. In school districts where Dr. Gordon's Jump Right In curriculum is used, this test may be a valuable assessment tool but music educators whose concept of musical aptitude and achievement goes beyond Edwin Gordon's rather narrow definitions will not be satisfied by this test. For example, tone production (instrumental or vocal), body-eye coordination, reading ability, and style sensitivity are just a few possible factors relating to music aptitude which are ignored by Dr. Gordon. Another problem is that although the revised version is simpler to apply and score, the normative data are based on the original 1970 national sample data (Radocy, 1998).

The Music Achievement Tests (Colwell, 1969; 1970b) assess achievement in a wide variety of listening activities for students from fourth to twelfth grade in a series of four tests. Teachers can select which tests best represent curricular objectives. Very extensive normative data are available based on grade and musical background although the normative data are all based on 1969 national data.

Probably the greatest weakness of the Iowa Tests of Music Literacy and Music Achievement Tests is that they are based on listening and although listening skills are required for effective performance, these tests do not directly measure performance

achievement. Since most secondary music education is performance based, effective assessment must directly examine performance.

The Watkins-Farnum Performance Scale (Watkins & Farnum, 1954, 1962) is used to examine sight-reading ability of instrumental music students. This test consists of a set of 14 exercises which is graded by taking away points for errors. The packet includes a detailed explanation of what constitutes an error. Sets are available for all band instruments and the Farnum String Scale (Farnum, 1969) offers essentially the same system for strings. Although Watkins and Farnum were concerned primarily with tonal and rhythmic errors, a Performance Rating Scale Supplement which deals with musicality aspects of performance was also developed. Although no normative data are available, correlations between the Watkins-Farnum Performance Scale and instructor ranking of those students are reported from .68 to .87 depending on the musical instrument. Despite its age and relative lack of supporting data, this test is often used in research as an objective performance measure (McPherson, 1994, 1995; Zdzinski, 1991, 1996).

Colwell (1970a) summed up the benefits of standardized testing for performance groups by the following statement:

In large performing groups where the age span may be as much as four years, the teacher requires norms by age, grade level, and type of instrument to determine the progress of individuals. Such norms are usually available only on standardized tests. (p. 17)

Although standardized tests may assist educators by providing instruments which have already been examined for validity and reliability, the variability and complexity of secondary instrumental program objectives make it unlikely that high school band

teachers will find standardized tests appropriate for all or most present objectives. If a test is found to appropriately measure one or more course objectives it must be determined if the time and expense of test administration are justified for the objectives measured. For example, if just one of eight or ten course objectives relates to sight reading, it must be determined if the time and expense of administering the Watkins-Farnum Performance Scale are justified for measurement of a single objective. Program assessment may be better served by standardized tests because measurements may be done less frequently than would be expected for traditional student assessment and the long-term normative data are more important.

#### Use of Technology in Music Assessment

Many modern technological advances may serve as valuable music assessment tools. The following is a brief overview of how some of these devices may be used to assist music educators with learning outcomes assessment.

#### Computer-based Sound Identification and Visual Representation

Since the early days of computers, musicians have seen the potential of visual sound representations for use in musical performance measurement. Freedman (1965) examined the use of an early computer to analyze tone quality. Although differences between good tone and poor tone could be demonstrated quantitatively, only single tones, rather than authentic musical performances, could be examined.

In 1969, the Computer-based Music Skills Assessment Project [CMSAP] began at the University of Illinois with funding from the National Endowment for the Arts. One of the project outcomes was development and evaluation of a computer system for measuring pitch and rhythmic accuracy of instrumental performance. This was done by

simply reporting actual versus expected frequencies (pitch) in vibrations per second and actual versus expected rhythms in microseconds. One of the difficulties with these early systems was that computer measurement (and reporting) of pitch and rhythm was too exact. Even virtuoso performers have slight but measurable pitch and rhythmic variations. Peters (1974) concluded that existing computer systems provided too little positive feedback and although these devices were shown to measure rhythmic and tuning with greater accuracy than human judges, they were not practical for regular use.

One of the reasons few music educators attempted to use computers for sound recognition and assessment in the 1970s and early 1980s is that only mainframe computers had sufficient storage capacity and processing speeds. However, by the end of the 1980s, desktop and portable computers with the capability to perform sound analysis were commonly available in schools. Although early sound analysis software was designed for speech and hearing use, music assessment and feedback was also possible (Zdzinski, 1991). Rees and Michelis (1991) used a 386 PC clone to examine performance imported through a Musical Instrument Digital Instrument [MIDI] port. Using Turtle Beach Sample Vision software, sound files were displayed as three-dimensional graphs. These graphic displays allowed for visual analysis of attack, dynamics, and overtone activity. In the Rees and Michelis study, time coded visual information was also collected using a video camera. Since the visual information was time coded (at 30 frames per second), comparisons between the visual images and graphic representation of the sound allowed for effective assessment of musical performance. This is the only study in which computer sound analysis was combined with time coded video recording.

Measuring rhythmic accuracy is perhaps one area where computers can be of greatest assistance.

Perhaps the most common method of scoring rhythmic performance is to listen to subjects' performances, deciding at the moment whether the performance is accurate . . . This method lacks objectivity because criteria are vague and the absence of a recorded copy prevents confirmation of scoring accuracy.

(Grieshaber, 1993, p. 128).

Rhythmic precision is relatively simple to measure using a computer since it does not always require sound recognition capability. In a study by Grieshaber (1993), musicians tapped on a device linked to a computer. Rather than just providing a visual representation of the performance, the program allowed distinctions between right hand and left hand performance, superimposition of many patterns on each other allowing for visual identification of variations on the same pattern. In addition, a visually displayed metronome was used to demonstrate precision problems. Taps were also evaluated and presented mathematically based on millisecond variation from the metronome and standard deviations for each. For example, it may be found that a student's taps are an average of 2 milliseconds different from the correct rhythm and 10% of the taps are more than 5 milliseconds off correct. The computation of this kind of numeric data in this study indicates greater potential for assessment and grading than did those procedures providing only visual representation.

In 1993, representatives of the Computer-based Music Skills Assessment Project announced the following findings:

1. Computer hardware and pitch-detection devices do exist for computer-based testing in music performance. 2. Pitch-detection hardware met the performance specifications established for successful evaluation of music performance. 3. Appropriate aural stimuli can be presented to inexperienced students to assess their ability to match pitches and to judge their tonal memory. 4. The CMSAP test instrument was deemed reliable, discriminating and appropriate for assessment of music performance of musically naive high school students. 5. Computer-based music performance testing was judged to receive a high level of acceptance by public school music administrators and computer technology administrators. 6. Computer software can be developed within a microcomputer environment to support computer-based music performance skills testing. (Peters, 1993, pp. 42-43)

Many easy-to-use computer software packages for analysis of musical performance are now available to music educators. For example, in a 1997 Instrumentalist article, Sound Explorer from Advantage Showare is presented as a way to help students "see the melodic contour of a phrase and the precise tuning of each note" ("What's New", p. 42). This program graphically displays vibrato, dynamics, accents, and articulation styles. It is likely that this program could also be used for assessment. This program is available for Windows or Macintosh and includes all the additional hardware (including microphone) needed.

It is interesting that in Performance Standards for Music: Strategies and Benchmarks for Assessing Progress Toward the National Standards (MENC, 1996), no

mention was made of the use of computers in performance assessment other than a brief reference to the use of computers and hand-held devices for the recording and compiling of grades. This rather universal technology is not developed specifically for music education but rather intended for use in all areas of education.

#### Audio/Video Recording of Student Performance

An important issue in musical performance assessment is reliability (MENC, 1996). In procedures where assessment is based on live student performance with a single rater, measurement accuracy and reliability cannot be evaluated. The use of audio or video recording "allows the scorer to better control the conditions under which the scoring is done and makes possible subsequent confirmation of the scoring if desired (p. 14)."

Although video is sometimes used for assessment of secondary music objectives, audio recording is more common for reasons of simplicity. Although some discussion regarding video recording versus audio recording may be found in the literature, no reliable evidence has been presented that indicates that one format is more reliable than the other. Colwell (1970a) wrote the following regarding audio and video recording of musical performance:

The video tape recorder offers all the advantages of the tape recorder plus the advantage of seeing visual causes of performance assets and defects. Embouchure, bow arm, hand position, and posture are all caught for the viewer to behold and evaluate . . . The great advantage of both devices is that they can repeat the identical process for several evaluators, so that the subjectivity of a single evaluator can be overcome. (p. 108)

This use of recording devices to control for a single evaluator's subjectivity is an important use not often mentioned in the literature, possibly because music educators are sufficiently busy evaluating their own students and it seems unlikely that music teachers would find time to evaluate recordings of students from other schools.

In a study by Davidson (1993), the relative importance of visual perception in musical performance assessment was examined. In two experiments, undergraduate music majors rated musical performance as "deadpan," "projected," or "exaggerated." In experiment one (N = 21), violin performance was rated and piano performance was rated in experiment two (N = 34). Raters were each provided with video, audio, and both video and audio of all performances in random order. Davidson offered the following conclusion:

The results suggest that vision can be more informative than sound in the perceiver's understanding of the performer's expressive intentions.

Indeed, in one experiment it was only vision mode that enabled the perceiver to discriminate between the three performance manners of deadpan, projected and exaggerated. (p. 112)

These findings may be a result of special conditions set up to enhance the visual perception such as tight-fitting black clothing and reflective tape affixed to the subjects' head, elbows, wrists, knees, ankles, hips, and shoulders. Findings may have little valid application to real musical assessment situations. Indeed, the relative merits of video should be studied in more authentic ways to determine the relative value of visual information in musical performance.



A possible problem associated with video rather than audio recording of student musical performance may be the special legal considerations associated with video. "Many schools have blanket policies regarding videotaping . . . [however, where such policies do not exist, it is still necessary] to obtain parental permission for videotaping" (Carlin, 1996, p. 39). Killian (1993) suggested the use of video within the rehearsal setting for assessment of specifically visual aspects of performance such as posture, hand position, mouth position, and breathing.

Carlin (1996) listed many possible uses of video, including documentation of student progress, creative process, inform teacher regarding classroom process, longitudinal assessment, and student and peer evaluation. Most of the ideas presented in this article were simply mentioned with no practical suggestions regarding implementation. A good motivational use of video was mentioned in this article that was not discussed in other sources. Carlin suggested videotaping initial sessions of a project for use in later rehearsals when students are tired and enthusiasm is low. "Students are cheered as they remember their initial excitement and creative output . . . [also] students can develop a sense of security in knowing that something artistic as been achieved" (p. 39).

In Performance standards for Music: Strategies and Benchmarks for Assessing Progress Toward the National Standards (MENC, 1996), both audio and video are suggested as ways to record student response:

Ideally, when the assessment strategy calls for the student to sing, play instruments, or move, the student's response should be audio taped or videotaped for subsequent scoring. That allows the scorer to better

control the conditions under which the scoring is done and makes possible subsequent confirmation of the scoring if desired. (p. 14)

Killian (1998) also suggested that assessment of sight-reading exercises and assignments be completed on an individual basis in another room using audiotape. In this way, students leave one at a time allowing rehearsals to proceed with minimal interruption but permitting periodic individual assessment of all students. Although assessment of performance in another room with a tape recorder is more authentic than other assessment strategies which do not directly measure performance, such assessment only measures performance outside of the ensemble setting. A method is suggested which allows for assessment of student performance within the ensemble performance setting:

That can be accomplished by using neck microphones and multiple tape recorders or a large multi-channel tape recorder. It may also be accomplished by using multiple small hand-held tape recorders or by having the teacher move around the room listening to each student.

(MENC, 1996, p. 14)

Other ideas for taping student performance within a rehearsal setting were put forth by

Killian (1998):

1. Tape individuals within the group by passing a small recorder from person to person (p. 11).
2. Tape the entire group with a single section . . . grouped around the microphone. You will be able to hear that section, but will also hear the rest of the organization; so the section will be heard in context (p. 12).
3. Tape the entire organization performing a selected piece

or prepare an accompaniment tape of selected portions of the piece.

Instruct individuals to then go to a quiet place and listen to the prepared tape while recording themselves performing their part with the tape.

(p. 13)

While the use of audio or video recording has long been a common idea in the music assessment literature, often the emphasis is placed more on the use of the information than the recording itself. For example, Rutkowski (1994) endorsed the use of audiotape but was more concerned with teachers understanding the purposes of evaluation as diagnostic, formative, or summative. Robinson (1995) discussed the use of audio or video recording as part of a portfolio kind of assessment but the emphasis was on rubrics for the evaluation of these recordings and the role of recordings in combination with other assessment data.

#### Performance Based Assessment in Music Education

As more and more emphasis is placed on authentic performance assessment, it is important that music educators use appropriate tools which address reliability concerns:

Demonstrations, projects, and portfolios offer a wide array of possibilities for authentic assessment. Because these forms require more than just answering questions . . . multilevel rubrics (tables, charts, or explanations of scoring categories or criteria) must be designed. Whichever type of assessment is implemented, the various levels of achievement must be predetermined by the music educator and understood by the class. In fact, students can be encouraged to assist in the development of the scoring rubric. (Cope, 1996, p. 41)

The use of good scoring rubrics improves reliability "both across different students and for the same student at different times" (Nitko, 1996, p. 261). Rubrics require several important steps: clearly define performance levels, craft performance tasks, and create scoring forms (Nitko, 1996). Although many rating scales and checklists are found in music assessment literature, they must be viewed within the context of these required steps for performance based assessment.

Rating scales often list specific aspects of student performance and a total number of points for each performance aspect, such as hand position - 10 points, posture - 15 points, dynamics - 10 points, and rhythmic accuracy - 20 points (Killian, 1998; Russo, 1988). Placing relative weighting of performance aspects does not meet the criterion of defining performance levels since there is no clear indication what constitutes eight points versus nine points for hand position. A person with a few minor hand irregularities could receive four points at one scoring and seven points at another due to this lack of specificity.

Matheny (1994) proposed a self-evaluation form for music ensemble students. In this form, students are asked to respond to ten items (attendance, effort, musical skill, technical skill, etc.) using ten point rating scales. For eight of the items, the top and bottom ratings are specified, and for two of the items, performance for several other scale points is also specified. Although the basic idea set forth in this article, use of student self-evaluation to clarify the teacher's grading, may be valid, the form shown in Figure 1 (p. 38) does not adequately specify performance levels to meet the criteria for an effective rubric.

McPherson (1993) developed a series of five point scales to measure improvisation ability among high school clarinet and trumpet students (N = 101). In each scale, performance descriptions for only one and five on the scale were specified. However, it was stated in the article that the three judges used in the study received training "to familiarize themselves with the evaluative criteria" (p. 15) so it is possible that judges were provided with clear indications regarding performance consistent with each point on the scale. Since reported inter-judge reliability ranged from .71 to .94, it is clear that judges used the scales consistently. It is likely that this inter-judge consistency was more a result of judge selection and training than the reliability of the rating scale.

Although scales described by writers in the field (Killian, 1998; Matheny, 1994; McPherson, 1993; Russo, 1988) are likely to improve the assessment of many music educators, they do not meet the criteria for valid assessment rubrics as described by Nitko (1996) because detailed description of performance levels was not provided. Many rating scales were found in the music assessment literature which meet the criterion of description of all performance levels (Cope, 1996; MENC, 1996; National Center for Education Statistics, 1998; Robinson, 1995; and Swanwick, 1994).

Swanwick (1994) clearly presents the use of rubrics in music assessment. An example of an eight level rubric for overall performance was included (pp. 108-109). Robinson (1995) presented ideas for collecting music performance information in a variety of formats, scoring rubrics, interviews, journals, and portfolios. A five level rhythmic scoring rubric was presented along with a checklist of performance qualities. Unfortunately, interviews, journals, and portfolios were presented as creative sources of student information but the use of rubrics with these other methods was not made

apparent. Cope (1996) briefly described the process of rubric development and presented two fine examples of five level music assessment rubrics, one rhythmic assessment rubric and one tonal assessment rubric. The form also presented a checklist for assessing musical expressiveness.

The rubrics set forth in Performance Standards for Music: Strategies and Benchmarks for Assessing Progress Toward the National Standards (MENC, 1996) define performance within three levels, basic, proficient, and advanced, at three age groups, Pre k-4, 5-8, and 9-12, as judged against nine content standards. All performance levels within this rubric are described in good detail, usually through the use of multiple descriptors. Assessment strategies are provided for all achievement standards to assist educators apply the rubric.

#### Actual Practices and Attitudes Toward These Practices

McCoy (1988) sent questionnaires to 396 randomly selected band and choir directors in the state of Ohio regarding grading criteria. Results indicated the following: (a) There was considerable variation in grading criteria; (b) what one director perceived as "A" performance was often perceived differently by other directors; (c) attendance and behavior were the most common non-music grading criteria; (d) 95% of the surveyed directors' grading systems included at least some non-music criteria; (e) 75% of directors' grading systems included at least some performance criteria; (f) 66% of directors' grading systems included at least some student attitude criteria; (g) 42% of directors' grading systems included at least some cognitive criteria; (h) directors perceived that school administrators placed greater emphasis on performance skills than non-music

criteria. Although this study was limited to Ohio and had only a 24% return rate, it provided the best data at the time regarding actual grading practices.

In a later study, McCoy (1991) sent surveys to principals, band directors, and choir directors at 98 randomly selected Illinois high schools. Completed questionnaires were received from 36 principals, 55 band directors, and 42 choir directors. Surveys included 25 possible criteria for determining grades divided into cognitive, psychomotor, affective, and non-music; participants were also encouraged to write in any other criteria not listed in the survey. Principals were asked what weight each criterion should receive in determining student grades. Directors were asked what weight each criterion received in actual grading practice. Additionally those directors and principals who included performance criterion in grading were asked whether performance should be evaluated against some fixed standard, other students or the director's perception of that student's potential.

Results of the study indicated that, in actual practice, non-music criteria were the most weighted criteria used to determine grades by band and choir directors. Concert attendance was the most weighted single item ( $M = 17.38$  for band directors and  $M = 14.72$ ) with the second heaviest weighted item being attitude for choir directors ( $M = 12.8$ ) and ability to perform concert music for band directors ( $M = 13.53$ ). Analysis of variance results indicated significant differences ( $p < .05$ ) between directors and principals in the weighting of non-music and cognitive criteria, with principals suggesting less weight for non-music criteria and greater weight for cognitive criteria than band and choir directors. When grading performance, directors and principals preferred to

compare student performance with directors' perception of student's potential rather than comparisons to a fixed standard or to other students.

Although this study was limited to Illinois, the findings confirm the 1988 McCoy study carried out in Ohio. The low return rate among principals may be the result of principals who are less familiar with music classes not returning the survey. This could bias the findings of principals. This study and the McCoy (1988) study were carried out before the assessment discussions which led to publication of the MENC standards. It is possible that changing attitudes toward music assessment may have caused changes which would invalidate these earlier studies.

Monroe (1995) compared the opinions of Ohio choir directors, band directors, principals, and college music education faculty regarding selected issues in high school music. Surveys were sent to the principal, choir director, and band director at 100 randomly selected high schools and 110 college instructors in music education of which a total of 234 usable surveys were returned (57% response). In one section of the survey, participants were asked to select the best description of actual practice and the best description of ideal from the following three general assessment descriptions: (a) There should be a specific course of study with measurable outcomes by which student learning can be determined; (b) While there should be general curricular guidelines, the program should be evaluated on performance results, such as quality of concerts, contests results, etc.; (c) As performance groups are in effect activities, program evaluation should be based on enrollment figures, student and parent satisfaction, and public reaction.

Most respondents in all participant groups (choir directors 64%, band directors 81%, principals 79%, university music teachers 86%) agreed that description "a" (specific



measurable outcomes) was ideal for high school music assessment. While band and choir directors were evenly divided between the three descriptions, description "a" (specific measurable outcomes) showed the highest mean response from principals (44.07%) and description "b" (general curricular guidelines) was perceived to be actual practice by most college music educators (61.70%).

Although actual practice was not measured in the same ways as in McCoy (1988, 1991), the finding that roughly one third of high school music programs include student assessment based on specific measurable outcomes is consistent with the findings of those studies. Differences of perspective between high school music teachers and principals regarding high school music assessment are also consistent with the McCoy (1988, 1991) studies. The differences between actual practice and ideal regarding assessment in high school music classes suggest that assessment may be changing and actual practice is not keeping pace with knowledge.

McClung (1997) examined attitudes toward assessment and grading practices in Georgia high school vocal music programs. Surveys were used to collect data from 615 Georgia Senior High All-State Choruses members (100% return rate), choir directors (80% return rate), and principals (78% return rate) from 150 schools with students in the Georgia Senior High All-State Choruses.

Respondents used a six point agree to disagree Likert scale to rate the suitability as grading criteria of the following: (a) sight-reading tests (b) on-the-music tests (c) pencil-and-paper tests (d) attendance (e) attitude (f) portfolios. The highest rated criteria by teachers and principals were sight-reading tests and on-the-music tests with 96% of teachers and 82% of principals rating both items as strongly agree or agree. The

highest rated criterion by students was attitude which was rated agree or strongly agree by 66% of the students. The lowest rated criterion among teachers was portfolios (35% agreement, 46% moderate, 19% disagreement) and pencil-and paper tests was the lowest rated criterion among principals (52% agreement, 53% moderate, 5% disagreement) and students (16% agreement, 53% moderate, 31% disagreement). All groups agreed (teachers 57%, principals 54%, students 54%) that six-weeks grades provide extrinsic motivation for students but in response to an item which asked if low grades impact group motivation, teachers and principals (this item was not present on student surveys) provided a bell-shaped response (21% agreement, 53% moderate, 24% moderate for teachers; 30% agreement, 45% moderate, 25% disagreement for principals. Relative use of grading criteria as perceived by students (only the student surveys addressed actual practice) was as follows: participation and attitude 84%, attendance 46%, individual performance assessment 35%, paper-and-pencil tests 8% (McClung, 1997).

McClung's findings confirmed those reported by Monroe (1995) in that while principals and choir directors' ideal assessment was based on measurable student performance outcomes. However, performance based assessment was not reflected in actual practice. One important difference between McClung's and Monroe's findings was the relative similarity between teachers' and principals' ideal assessment. In the Monroe study principals' response differed significantly from choir directors in that a greater proportion of principals endorsed assessment based on specific measurable outcomes but this may be the result of sampling. The use of an all-state choir as the basis for the sample may result in serious external validity problems since programs with

students in the all-state choir are likely to differ in many ways from other vocal programs whose students are not in the all-state choir. Unfortunately, actual practice was only examined based on student perception. Actual practice as reported by teachers may differ from student perceptions. Although band programs were not examined in this study based on the findings of McCoy (1988, 1991) and Monroe (1995) it is reasonable to conjecture that the opinions of band directors are similar to those of the choir directors.

#### Definition of Assessment in Music Education

The definition of what constitutes assessment is disputed in the education community. Many music educators hold to the traditional perspectives of assessment, which define assessment as being "a formal appraisal of the quality of educational phenomena" (Popham, 1993, p. 7). Hoffer (1993) described the process as follows:

Assessment of what students have learned in a music class or rehearsal is the other side of the coin from planning. The two aspects of teaching are, or should be, that closely related. In fact, assessment is not even possible unless the objectives have been clearly stated. (p.29)

Lehman (1992) questions the notion that it is "possible to assess a student's performance without hearing him or her alone" (p. 58). Given the conservative perspective of traditional assessment, music education, especially high school performance ensembles, appears to lack necessary learning assessment.

Many music educators endorse a wider perspective regarding assessment. "To most music teachers, the ultimate test of a performing group is how it sounds. Any other evaluation would be superfluous and a waste of time" (Lehman, 1992, p. 57). Music educators endorsing this point of view hold that some form of assessment is inevitable in

the music education process. Almost every word spoken by a conductor while rehearsing with any ensemble is based upon assessment. For example, if a conductor asks clarinets to use softer attacks in some section or reprimands brass players for not allowing a delicate melody to come through, those comments (feedback) are based upon assessment of previous performance (Swanwick, 1994). Given a more liberal definition, rehearsals, rather than being devoid of assessment, are made up almost entirely of assessment and feedback.

#### Roles of Assessment

Hoffer (1993) outlined the following reasons for assessing: (a) It provides information for good teaching, (b) it provides evidence of learning to educational agencies and school boards, (c) it can lead to more valid grading. Lehman (1992) wrote, in a journal for secondary school principals, "that thinking carefully about student assessment forces teachers to think carefully about their objectives" (p. 57). This was demonstrated by Graham (1989) who examined music programs in Canada. He found that music programs that included both large-scale and small-scale assessments had more appropriate and more clearly stated curricular objectives.

Unfortunately, many music educators see assessment as simply a requirement to ensuring a place for music in the curriculum. This view of assessment, rather than being motivated from a desire to improve music education, is a reaction to statements like "What gets tested, gets taught; what isn't tested, isn't taught" and "What is important is tested and what is tested is important" (American Council for the Arts, 1995, p. x). This rationale is discounted as unworthy because "testing in music should be done for the best reasons that we test in other disciplines" (Lehman, 1992, p. 57).

### Summary

A wide variety of equipment, standardized tests and strategies is available to music educators. Software has been developed to allow for visual analysis of musical performance (sound). Standardized tests are available to assist music educators with assessment of musical performance, musical knowledge, and musical aptitude. Many strategies for use of audio and video tape recording for assessment are readily available in music education journals. Rubrics have been developed and validated to assist music educators in carrying out valid and reliable assessment.

Subsequent chapters describe how data were collected and analyzed to determine the current use or non-use of available assessment methods in music education, the use of assessment in grading, satisfaction with current practice, and factors relating to current practice.

## CHAPTER III

### METHOD

#### Introduction

In this study, I examined current assessment and grading practices in United States high school bands and local satisfaction with those practices from the perspective of band directors. Procedures used in this study followed the general guidelines outlined for descriptive research in music education (Casey, 1992; Phelps, Ferrara, & Goolsby, 1993).

This investigation required the development and administration of appropriate data gathering instruments and selection of appropriate participants. The present chapter includes descriptions of the following phases of the data gathering process: selection of participants, development of surveys, procedures for the pilot study, procedures for the primary study, and a description of data analysis procedures. A separate section is devoted to each topic.

#### Selection of Participants

High school band directors were drawn from selected schools. A total of 600 public high schools were selected using stratified random sampling among the six MENC geographic regions. Selection was completed by compiling a list of public high schools (including addresses) within each MENC region using online school directories, primarily School Match ([www.schoolmatch.com](http://www.schoolmatch.com)) and American School Directory

(www.asd.com). Only public high schools were included in the regional databases. Very small schools (schools with either reported enrollments of less than 50 or less than 3 teachers) and specialized schools (e.g. special education schools, juvenile detention schools) were deleted from the regional databases because these schools were unlikely to have band programs which would provide usable data for this study. Two small, rural K-12 schools (with high school enrollments of less than 50 students) had inaccurately reported their high school enrollment figures and were subsequently included in the study. Schools were then selected from each regional list using a computerized random-number generator (SPSS Inc., 1999). Initially, 75 schools from each MENC region were selected. Selection of another 150 schools was carried out with the number of schools from each region determined by the proportion of US high schools located in that region. MENC regions, states within each region, the proportion of United States high schools found within each region, and the number of schools selected from each region are shown in Table 1.

Surveys (Appendix A) and cover letters (Appendix B for the first mailing, Appendix C for the second mailing) were mailed to the band director of each school (after being approved by the University of North Dakota Institutional Review Board). In schools with more than one band, and/or more than one band conductor, data were collected from the band with the most eleventh grade students. This was done to simplify data collection and avoid bias that could result from varying assessment and grading practices within participating schools. It was thought that whichever band contained the most eleventh grade students would better represent an overall system because eleventh grade students would not likely receive special grade level related treatment or privileges.

For example, freshman band participants may have more individual attention than older members because of their youth. A senior class band may receive special privileges or experience other unique circumstances.

Table 1

MENC Regions, States Included in Regions, Percentage of US High Schools by Region, and Number of Selected Schools by MENC Region

Region	States	N <sup>1</sup> Schools	% <sup>2</sup> National	N <sup>3</sup> Sample	% <sup>4</sup> Sample
Eastern	Connecticut, Delaware, Washington DC, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont	2832	15	98	16.3
Southern	Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia	2687	18	102	17.0
North Central	Illinois, Indiana, Iowa, Michigan, Minnesota, Nebraska, North Dakota, Ohio, South Dakota Wisconsin	4023	26	113	18.8
South Western	Arkansas, Colorado, Kansas, Missouri, New Mexico, Oklahoma, Texas	3047	20	105	17.5
North Western	Alaska, Idaho, Montana, Oregon, Washington, Wyoming	1115	8	87	14.5
Western	Arizona, California, Hawaii, Nevada, Utah	<u>1856</u>	<u>13</u>	<u>95</u>	<u>15.5</u>
Total		15560	100	600	100

<sup>1</sup> Number of schools in the region

<sup>2</sup> Percentage of US schools in the region

<sup>3</sup> Number of schools sampled in the region

<sup>4</sup> Percentage of the sample in the region



### Development of Surveys

A survey was developed to collect data from high school band directors (Appendix A). The survey was pilot tested using small and convenient samples near the University of North Dakota. The entire survey is shown in Appendix A with item numbers added for ease of communication. These were not present in the original survey.

For the pilot test, 5 high school band directors (3 from large suburban high schools and 2 from small rural schools) completed surveys, provided written evaluations of the surveys, and were interviewed regarding survey format, item clarity (wording), and survey content. All appropriate changes and improvements indicated by the pilot study were made before initiating the primary study. Most of the changes were made in the demographic information area where readers were sometimes uncertain whether questions referred to the citywide, school, or teacher-specific information. In the primary study, an online survey identical to the paper-and-pencil survey was also used for data collection ([http://www.members.tripod.com/Simanton/music\\_survey.html](http://www.members.tripod.com/Simanton/music_survey.html)).

### Procedures

The initial mailing was sent out April 21, 2000, and a second mailing was sent on May 15, 2000. The business reply envelopes used in the first mailing were marked with code numbers to identify respondents so that the second mailing would only be sent to those who had not responded to the initial mailing. Wherever possible, participants were also contacted electronically and offered the option of completing the online survey. The online survey address was also included in the cover letter of the second mailing. Data

were entered into the computer and analyzed using SPSS for Windows, Version 10.0.5 (SPSS Inc., 1999).

### Analysis of the Data

Descriptive statistics (means and frequencies) were used for the demographic information. Comparisons between current practice grade weights and optimal grade weighting were completed using paired sample t-tests.

### Regional Weighting

Assessment, grading, and satisfaction questions were answered using descriptive statistics that have been regionally weighted to ensure that data from each region was weighted appropriately to represent that region's proportion of US public high schools. For example, 20% of US public high schools are found in the states comprising the MENC southwestern region. Respondents in this study from that region make up only 16.8% of the respondents in this study. By slightly increasing the weight of respondents from the southwestern region, the voice of southwestern band directors is appropriately included in the calculation of national trends.

### Examination of Group Differences

Because of the large number of demographic and dependent variables, comparison of all dependent variables by all demographic variables was impractical. Therefore, a more manageable number of demographic and dependent variables was selected. The process for completing this is described below.

### Selection of Independent Variables

A preliminary examination of the seven demographic variables listed in research question four (differences in responses based on demographic variables) was carried out

through the use of three multiple analyses of variance (MANOVA), one each for regional factors (MENC regions and rural versus urban), school factors (school size and band size), and band director factors (educational background, years experience, and major instrument). The dependent variables in each calculation consisted of the 5 current grading practice variables (Appendix A, items 61-65) and the eleven opinion items (Appendix A, items 75-85). Variables with significant Pillai's trace findings in these initial analyses were included in further analysis of group differences. In other words, independent variables which produced no significant findings were dropped from subsequent analyses.

#### Selection of Dependent Variables

Rather than calculating differences with all variables combined, variables were grouped for analysis based on the initial research questions, assessment variables, grading variables, and assessment. Within each section (for example, assessment), only the primary indicator variables were used in order to improve the family-wide error rate and to avoid large numbers of highly correlated variables. For example, of the five tape recorder use variables (Appendix A, items 33-37), only the use versus non-use variable (Appendix A, item 33) was included. The other four tape recorder variables dealing with types of use (and frequency for each) were not included in analyses.

#### Examination of Differences

Relationships between selected demographic (independent) and dichotomous dependent variables were examined using cross-tabulations and Chi-square tests. The family-wide error rate was controlled through adjustment of the alpha level by dividing the usual alpha (.05) by the number of tests minus one (Freedman, Pisani, & Purves,

1980). For example, in the first section (assessment variables) where four dichotomous variables were examined using Chi-square the alpha was adjusted to .017 (.05/3).

Relationships between selected demographic (independent) and continuous dependent variables were first examined using multiple analyses of variance (MANOVA). Demographic variables found to be significantly related (using Pillai's trace) were then examined using one way ANOVAs with family-wide error rate controlled by Bonferoni's inequality. For independent demographic variables (with significant relationships) made up of more than two groups, final analysis was carried out with a series of post hoc Bonferoni t-tests.

### Summary

Chapter III describes selection of subjects, instrument development, and data analyses employed to address the research questions posed in Chapter I. Data collection results and findings with regard to the research questions are presented in Chapter IV.

## CHAPTER IV

### RESULTS

In this investigation, I examined current assessment and grading practices in United States high school bands and local satisfaction with those practices. In this chapter, I will present the results of data analyses as follows: (a) survey response data, (b) demographic information, (c) summary of assessment variables, (d) summary of grading variables, (e) summary of satisfaction items, and (f) relationships between demographic variables and assessment, grading, and satisfaction variables.

For calculation of overall findings, data were regionally weighted to more accurately represent national trends. This was done because the proportion of respondents from each region did not match the predicted value, that is the proportion of high schools in each region. Regional weighting causes the proportion of data from each region used in calculations to be the same as the proportion of US high schools found in that region. All weightings used were between .5 and 1.25. Regional weighting was used to calculate findings in the following sections: (a) demographic information, (b) summary of assessment variables, (c) summary of grading variables, (d) summary of satisfaction items (SPSS Inc., 1999).

#### Survey Response Data

Of the 600 surveys sent, 27 participants responded electronically and 175 responded to the paper-and-pencil surveys, for a total of 202 responses. As shown in

Table 2, regional response rates varied from 31.4% in the southern region to 38.1% in the north central region.

Table 2

## Survey Return Information by Region

MENC Region	Sent	Returned	Return Rate %
Eastern	98	31	31.6
Southern	102	32	31.4
North Central	113	43	38.1
South Western	105	34	32.4
Western	95	31	32.6
North Western	87	31	35.6
Total	600	202	33.7

Of the 600 surveys, 9 were returned as either undeliverable or the schools reported having no band program. Taking these surveys into account, the corrected response rate was 34.2%.

Online responses were compared to traditional survey responses using a one way ANOVA with response format being the factor and the current grading system variables (Appendix A, items 61-65) and the 11 opinion variables (Appendix A, items 75-85) as dependent variables. No relationships were noted between response format and any of the examined variables. For subsequent analysis, online surveys are included with the paper-and-pencil ones.

### Demographic Information

Information was collected regarding the following: (a) education and background of respondents, (b) school setting information such as population, socio-economic status, school enrollment, etc., (c) band program information including number of bands, number of other ensembles, number of teachers, and percentage of students who participate, etc., and (d) band information regarding the specific band being examined. For more information about the weighting procedures, see Chapter III.

#### Education and Background

All survey participants reported having earned at least a bachelor's degree. Almost half (48.9%) of the participants also reported having completed a master's degree and two participants (.8%) reported having earned doctorate degrees. Most of the participants' bachelor's degrees were in either music or music education (92.0%) as were most participants' master's degrees (93.3%).

Most band directors were brass players (60.8%); 28.4% reported a major instrument in the woodwind family and 10.8% reported majoring in percussion, strings, or other instruments. Participants reported from less than 1 year experience up to 40 years experience as band directors with a mean of 15.24 (SD = 10.44). They also reported having been in their current position for an average of 8.97 (SD = 8.83) years.

#### School Setting

Enrollments in schools included in the study ranged from 31 to 3,200 students (Median = 800) with a mean enrollment of 922 (SD = 721.29). These schools were located in towns or cities ranging in population from 100 to 17,000,000 with a median town population of 8,835. Most participants (60.1%) described the socio-economic status

of the majority of their students as “middle” while 30.3% described most of their students as lower socio-economic status and 9.6% reported that most of their students were of high socio-economic status.

#### Band Program Information

The number of concert bands per high school ranged from 1 to 5 with a mean of 1.88 (SD = 1.04) bands. The number of other ensembles (including jazz bands, pep bands, quintets, trios) ranged from 0 to 20 with a mean of 2.27 (SD = 3.02) ensembles per school. The percentage of high school students participating in the band program ranged from 1% to 88% of the total high school population at each school with a mean participation of 17.04% (SD = 15.71).

The number of band teachers ranged from .3 to 4.5. These teachers taught from 1 to 5 bands with a mean of 2.73 (SD = 1.33). It was noted that the mean number of bands taught by participating band directors exceeded the number of high school bands in the school. It is likely that this discrepancy occurred because many band directors also taught elementary or middle school bands either in the same building or nearby. Participants also taught from 0 to 14 other ensembles or classes (M = 2.41, SD = 2.25)

#### Band Information

Bands described by the study ranged from 10 to 200 members with a mean of 60.88 (SD = 31.01; median = 56) band students. The players in the bands examined in this study had a mean of 5.40 (SD = 1.49) years band experience. An average of 20.91 (SD = 15.96) players in each band (or roughly one third of all band students described in this study) are members of other instrumental ensembles.



Most bands described in this study (68.5%) included students from all high school grade levels and another 7.7% of the bands also included junior high grades. Just two of the bands described in this study (.9%) were made up of only eleventh grade students.

#### Summary of Assessment Data (Research Question One)

What do high school band directors report doing (what strategies and how frequently) to assess student learning within their band programs? In the first part of this section, assessment of student instrumental performance is examined. In the second part of this section, other indicators of student learning are investigated.

#### Assessment of Student Instrumental Performance

Frequency of student performance assessment is shown in Table 3. Note that while slightly more than 30 % of band directors do not assess student performance during rehearsals and 35.4 % of band directors do not assess student performance outside rehearsal time, 18 % of band directors report no assessment of individual student performance. It was also found that 32.2 % of band directors report that they assess each student's performance (either within rehearsal or outside of rehearsal) more than once per week. Roughly 1 of every 6 band directors reported no assessment either within or outside of rehearsals.

Among those band directors who reported assessing student performance, the length of student performance assessments ranged from 3 seconds to 20 minutes with a median assessment duration of 60 seconds. A more detailed description of assessment duration is shown in Figure 1.

Table 3

## Frequency of Individual Student Performance Assessment

Frequency of Assessment	Within Rehearsal <sup>1</sup>	Outside of Rehearsal <sup>1</sup>
None	30.1	34.2
Once/month or Less	27.6	28.8
2-3 Times/ Month	17.1	20.2
Weekly	14.4	10.4
More than Once/Week	10.8	6.4

<sup>1</sup> Reported as percentage of band directors

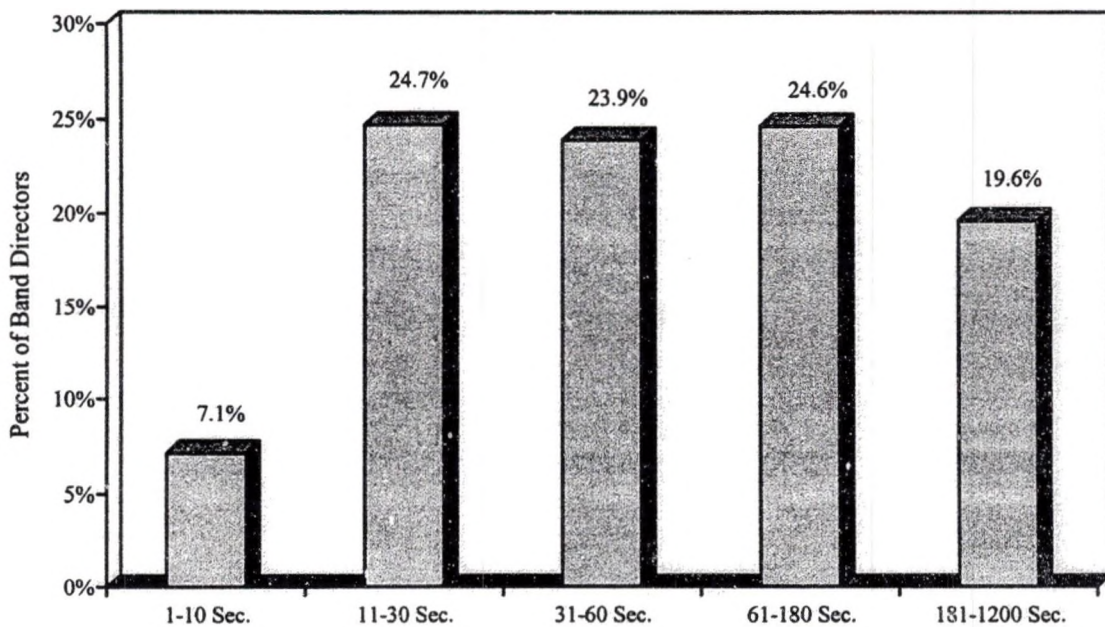


Figure 1. Distribution of student assessment duration categories in seconds.

Table 4 shows the percentages of band directors reporting use of tape recorders, video cameras, or computers for assessment as well as the mean frequency per month. Tape recorders are used by more band directors and more frequently than video cameras. Rehearsals are more often recorded (by tape recorders and/or video cameras) than individual students. Two participants (1.3%) reported using video portfolios for all students to illustrate performance improvement.

Table 4

Percentage of Band Directors Who Audio Tape and/or Video Record Student Performance

Recording Equipment Use	Percent Usage	Frequency Per Month	
		Mean	SD
Tape Recorder for Rehearsals	47.9	2.67	2.57
Tape Recorder for Individuals	33.2	1.18	1.21
Video Camera for Rehearsals	26.1	1.20	1.41
Video Camera for Individuals	6.7	.24	.43

Although computer usage for student learning assessment was reported by 6.6% of band directors, none of the software that band directors reported using is intended for performance assessment. Frequency data were not collected for use of computer for performance assessment.

Assessment of Non-Instrumental Performance Student Learning

Participants reported use of several assessment strategies to measure student learning other than instrumental performance. The most common of these strategies are shown in Figure 2. The most commonly used assessment shown in Figure 2 was quizzes, used by 41.7% of band directors. Although it was expected that the non-performance assessment would be used less than performance assessment, more band directors employed journals (13.2%) and theory exams (8.7%) than video taped individual student performance (7.6%).

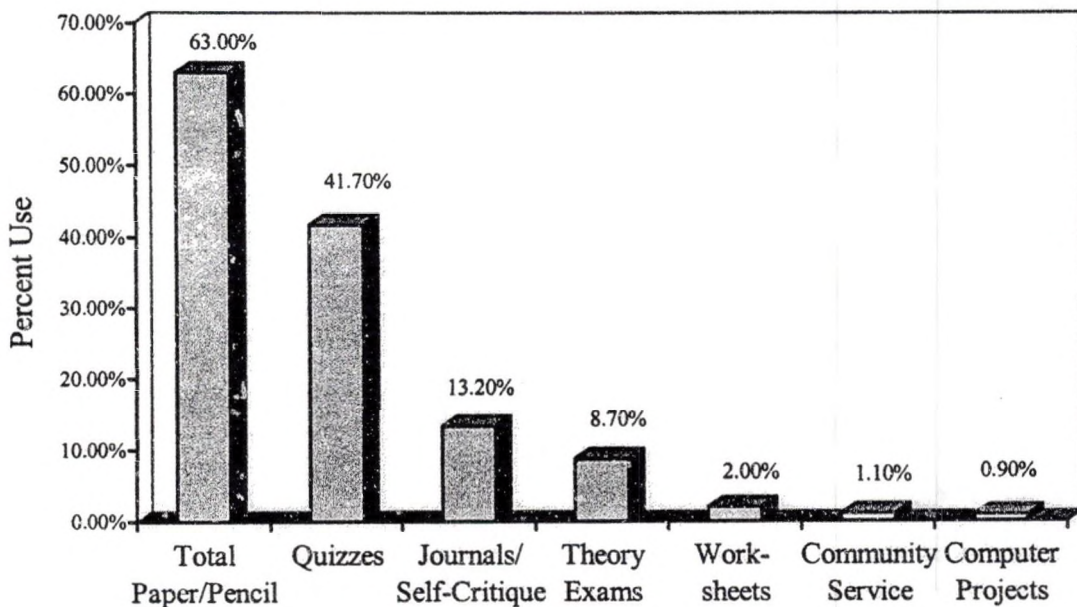


Figure 2. Non-instrumental performance assessment use (percentage of band directors).

Summary of Grading Data (Research Question Two)

Research question two reads as follows: What factors are reported to be used (and in what percentages) in high school band grade assignment? Findings for grading practices have been regionally weighted to more accurately represent national trends.

Participating band directors were asked what percentage of band students' grade currently came from each of the following criteria: (a) attendance, (b) participation/attitude, (c) performance of band music, (d) technique and/or sight reading, (e) other. Results are summarized in Table 5.

The high standard deviations in Table 5 suggest wide variation in grading systems; as a result, the raw frequency distribution is shown in Table 6. Relatively few band directors report basing more than 50% of students' grades on a single criterion. While only 2.8% of band directors report not using participation/attitude in grading, 42.6% report that they do include technique in their grading of band students. "Other" responses were made up primarily of practice logs and written homework/tests.

Table 5

Mean Percentage of Grading Criterion used in Student Grading

Grading Criterion	Percentage of Student Grade	
	Mean	SD
Attendance	25.7	19.5
Participation/Attitude	30.3	15.6
Band Music Performance	25.9	16.4
Technique/Sight Reading	10.6	12.6
Other (Non-Performance)	7.5	12.6
Total	100.0	

Note. Participants were required to subdivide grading percentages so that they summed to 100.

Table 6  
Percentage of Grading Criterion Weighting

Criteria	Not Used in Grading	1-25% of Grade	26-50% of Grade	51-100% of Grade
Attendance	17.1	41.8	35.1	6.0
Participation	2.8	47.0	44.8	5.4
Band Music Performance	12.7	47.5	35.9	4.0
Technique	42.6	47.2	9.5	.6
Other (Non-Performance)	67.1	23.9	8.5	.5

Note. Numbers shown in table represent percentage of band directors

Table 7 shows the proportion of band directors who inform students, parents and principals regarding grading policies. Students were reported to be most frequently informed regarding band grading policies (97.3%) and the principals were least frequently informed. Just over one third (36.2%) of band directors reported verbally informing their principals of band grading systems. The percentage of band directors who informed their principals regarding band grading systems varied significantly by geographic region which will be discussed later.

Band grade distribution is shown in Figure 3. Although most band students get "A's" in band, participants report that roughly 25% of their students receive band grades of "B" or less.

Table 7

Percentage of Band Directors Who Inform Students, Parents and Principals Regarding Band Grading Systems

	Informed <sup>1</sup>	Verbally <sup>1</sup>	In Writing <sup>1</sup>
Students	97.3	70.4	90.2
Parents	90.4	42.0	88.1
Principal	81.0	36.2	77.5

<sup>1</sup> Reported as percentage of band directors

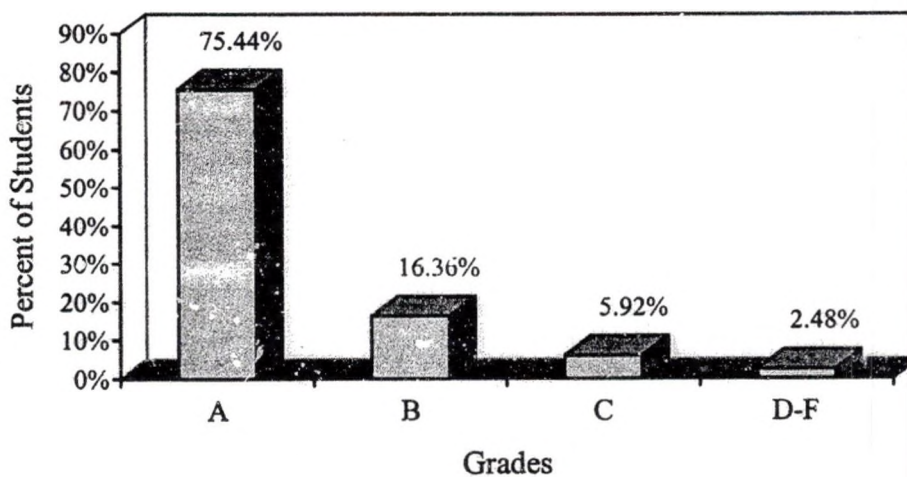


Figure 3. Reported grade distribution for band students.

#### Summary of Assessment and Grading Satisfaction Data (Research Question Three)

Question three was listed as follows: To what degree are the high school band assessment and grading practices viewed as adequate and appropriate from the band

director's perspective? In addition to providing information regarding current practice, participants were asked to provide data regarding the proportion of the student grades that should be based on each of the provided criteria (Appendix A, items 70-74). Information regarding participants' thoughts about the role of each criterion in grading and difference from current practice (see Table 5) is entered in Table 8. Results show a significantly lower mean for attendance and participation and significantly higher means for band music performance, and technique.

Table 8

## Comparison of Current and "Should Be" Grading Criterion Weighting

Criterion	Current		Should Be		Difference	t	2 Tail Sig.
	Mean	SD	Mean	SD			
Attendance	25.7	19.5	20.0	14.93	-5.7	5.499	<.001
Participation	30.3	15.6	28.0	13.42	-2.3	2.536	.012
Band Music Performance	25.9	16.4	28.8	14.67	+2.9	-3.006	.003
Technique	10.6	12.6	16.0	12.67	+5.4	-6.539	<.001
Other	7.5	12.6	7.2	12.27	-.3	.387	.699

Participants responded to 11 assessment/grading statements on a five point Likert scale from strongly disagree (1) to strongly agree (5). Table 9 shows mean results in descending order (stronger agreement levels first). The statement "I would do more student learning assessment if I had more time" garnered the strongest agreement



( $M = 1.55$ ,  $SD = .79$ ). The only items with mean responses on the disagree side of the scale (below 3.00) were “I am concerned primarily with ensemble rather than individual assessment” ( $M = 2.99$ ,  $SD = 1.68$ ) and “I think it is unfair to grade students by how well they play” ( $M = 2.47$ ,  $SD = 1.07$ ).

#### Relationships Between Demographic and Dependent Variables (Research Question Four)

Question four was stated in the following words: Which of the following factors may be associated with variations in reported assessment and grading practices and/or perspectives regarding those practices: regional factors (MENC region, urban versus rural), school factors (school size, number of students per band), band director factors (educational background, years experience, major instrument)?

The preliminary analysis described previously indicated few significant differences between levels of independent variables on the 15 dependent variables selected for analysis. The exceptions were as follows: (a) MENC region, (b) band size (small band = less than 60 members, large band = 60 members or more), (c) years experience categories, and (d) bachelor's degree versus graduate degree. As would be expected, years experience and possession of a master's degree were highly related ( $F = 44.248$ ,  $p < .001$ ) and the relationships between these two variables and the dependent variables were similar. Rather than carrying out calculations using two highly correlated variables (which could bias results because of multicollinearity), years experience was not included in group differences calculations.

Table 9

Mean Response to Assessment and Grading Opinion Items in Descending Mean Order

Statement	Mean	SD
I would do more student learning assessment if I had more time.	4.42	.79
School officials in my school seemed to be satisfied with current band assessment practices	4.23	.71
School officials in my school seemed to be satisfied with current band grading practices.	4.22	.62
Parents in my school seemed to be satisfied with current band assessment practices.	4.14	.64
Parents in my school seemed to be satisfied with current band grading practices.	4.10	.60
Band students in my school seem to be satisfied with current band grading practices.	4.07	.67
Band students in my school seemed to be satisfied with current band assessment practices.	4.02	.71
My assessment methods are good enough to ensure quality instruction.	3.76	1.68
My assessment and grading practices are similar to those of most of the band directors I know.	3.50	.87
I am concerned primarily with an ensemble rather than individual assessment.	2.99	1.68
I think it is unfair to grade students by how well they play.	2.47	1.07

Note. Based on the following scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree

#### Differences Between Levels of Education, Band Size, and Region

Differences were examined between levels of demographic independent variables for dependent variables using Chi-square (for nominal dependent variables) and

multivariate analysis of variance (MANOVA) (for continuous dependent variables). Rather than calculating differences with all variables combined, variables were grouped for analysis based on the initial research questions, assessment variables (4 dichotomous and 3 continuous variables), grading variables (3 nominal dependent variables and 9 continuous dependent variables), and assessment (16 continuous variables).

In the Chi-square calculations, the alpha level was adjusted to control for the family-wide error within each section by dividing the usual alpha (.05) by the number of dependent variables being examined minus one. For example, in the first section (assessment variables), where four dichotomous variables were examined, the alpha was adjusted to .017 (.05/3). Significant MANOVA findings were further examined (on a post hoc basis) through one way ANOVAs with family-wide error rate controlled by Bonferoni's inequality. For demographic variables (with significant relationships) made up of more than two groups (MENC region, levels = 6), final analysis was carried out with a series of post hoc Bonferoni t-tests.

In the following three sections, assessment, grading practices, and satisfaction with assessment and grading are separately addressed. Within the first two sections, subsections are devoted to nominal and continuous dependent variables. In the third section, satisfaction, only continuous variables are used in the analysis.

#### Assessment Variables

The assessment variables can be divided into two main sections, selected equipment used variables (dichotomous data) and assessment frequency/duration variables (continuous data). Among the "equipment used" variables, only the primary (use or non-use) variables were included while more detailed type of use and frequency

variables were not included. For example, under use of tape recorder, the use versus non-use variable was included in the analysis while the variables that specify rehearsal or individual use, and the frequency of those uses, were not included. This was done to simplify analysis and because the detailed response variables by necessity were highly correlated to the use versus non-use variables.

### Nominal Assessment Variables

Nominal assessment practices variables were examined using cross-tabulations and Chi-square tests of independence. The significance level was adjusted to .017 (.05/3) in order to control for examination of four dependent variables. Use of tape recorder for assessment (dichotomous: use versus non-use) was found to be significantly related to band size ( $\chi^2 = 8.183$ ,  $p = .004$ ) and band director master's degree ( $\chi^2 = 5.770$ ,  $p = .016$ ). Tape recorder use differences by band size and director education are shown in Figure 4.

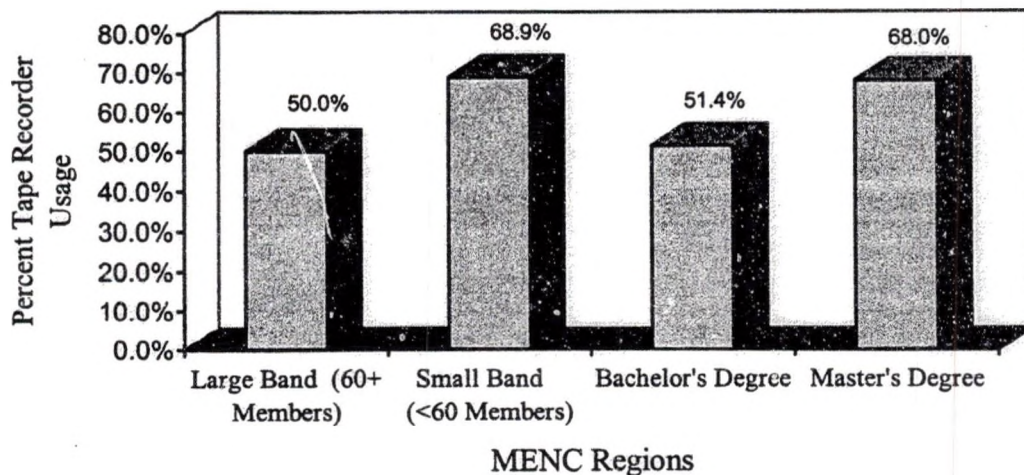


Figure 4. Percentage use of tape recorder based on band size and band director education.

A possible interpretation of these findings could be that band directors with only bachelor's degrees tend to conduct larger bands than those with master's degrees. The opposite was found to be the case. Significantly more band directors with master's degrees were found to conduct large bands than those with only bachelor's degrees.

#### Continuous Assessment Variables

A three way MANOVA was run on three continuous assessment practices variables (Appendix A, items 30-32). A significant Pillai's trace ( $F_{15,489} = 1.731$ ,  $p = .042$ ) for MENC region was noted and post hoc one way ANOVAs indicated significant differences for the dependent variable "assessment duration" between MENC regions ( $F_{5,181} = 3.937$ ,  $p = .002$ ). Bonferoni post hoc t-tests indicated the existence of significant differences between assessment duration in the north central region versus the western, south western, and north western regions. East and south regions did not differ from one another nor from any of the other regions. Mean assessment duration by region is shown in Figure 5. With a slightly larger sample size, given the trend shown in Figure 5, it is likely that two groups would evolve, east and north central versus the other four regions.

#### Grading Variables

Variables relating to grading policy include reporting variables which indicate who was informed of grading policies (Items 52, 55, 58, Appendix A; nominal data) and percentages of current practice grade weighting (continuous data). Among the reporting variables, only the primary variables (who was informed of grading policies, students, parents, and/or principal) were used while the reporting method variables (verbally or in writing) were not included in an effort to improve the family-wide error rate.

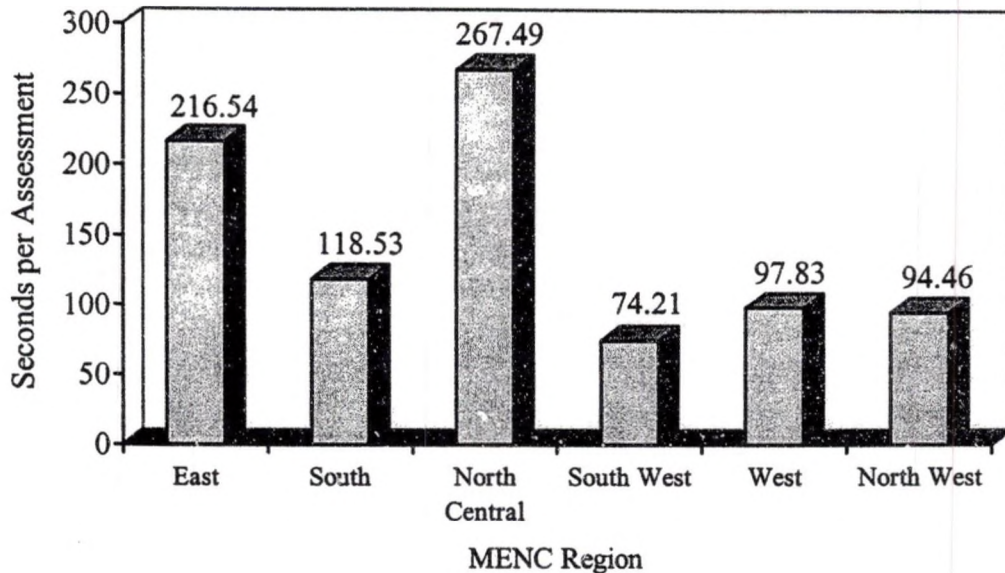


Figure 5. Mean assessment duration by MENC region (dependent variable expressed in seconds of assessment).

### Nominal Grading Variables

No significant differences were found between the MENC regions, band size, or band director education levels (master's degree) regarding who was informed of grading policies. In other words, none of the independent variables predicted practices regarding who gets informed.

### Continuous Grading Variables

The only significant difference found among the grading variables was a two way interaction effect, MENC region by band size (Pillai's trace  $F_{25,875} = 2.308$ ,  $p < .001$ ) In subsequent one way ANOVAs, significant relationships were found for region by band size for the following dependent variables: participation, band music performance, and "other" grade weightings.

Significant post hoc Bonferoni t-test results show the following three significant findings: (a) Within the MENC western region, directors of small bands weight participation significantly less (22.93% of total grade) than did directors of large bands (57.14% of total grade); (b) in the north central region, directors of small bands weight band music performance significantly less (14.81% of total grade) than directors of large bands (29.75% of total grade); and (c) in the eastern region, directors of small bands weighed other (paper-and-pencil tests and practice logs) significantly more (22.75 of total grade) than did directors of large bands (6.05% of total grade).

#### Satisfaction Variables

The variables used to determine satisfaction were the 11 opinion variables (using a five point Likert scale) and a series of variables calculated from the current practice grading variables and “should be” grading variables. These grading satisfaction variables were calculated by subtracting the current practice variables from the corresponding “should be” variables. All variables examined in this section were continuous. No significant differences were noted in assessment and grading satisfaction based on band size, MENC region, or band director education.

## CHAPTER V

### CONCLUSIONS

#### Introduction

The purpose of study was to examine current assessment and grading practices in United States high school bands and local satisfaction with those practices as reported by band directors. A survey was developed to collect demographic information, assessment data, grading data, and satisfaction data. Databases were created containing names and addresses of all public high schools within each of the six MENC regions. Six hundred schools were selected from databases and surveys were sent to the band directors of the selected schools (591 were delivered). An online survey was also provided for participants who preferred to respond electronically. Two hundred two usable surveys were returned for a final response rate of 34.2%.

#### Summary of Findings Regarding Research Questions

##### Question One: What do High School Band Directors Report Doing (What Strategies and How Frequently) to Assess Student Learning Within Their Band Programs?

Nearly 70% (69.9%) of band directors perform student assessment during rehearsals, and 65.8% report assessing student learning outside of rehearsals. More than 1 in 6 band directors (17.2%) report doing no individual student learning assessment whatsoever. Only about one third of band directors (33.2%) report using audio tape to record individual student performance and roughly 1 in 15 (6.7%) report doing so with a



video camera. Sixty-three percent of participating band directors report using paper-and-pencil for assessment for such things as quizzes (41.7%) and journals (13.2%).

In summary, it appears that while most band directors attempt some kind of individual learning assessment, (82.8% assess performance, 41.7% use paper-and-pencil) relatively few use recording equipment necessary for thorough, reliable assessment.

Question two: What Factors are Reported to be Used (and in What Percentages) in High School Band Grade Assignment?

On average, 56% of band grades come from non-performance criteria (attendance, participation, and attitude). Performance of band music accounts for another 25.9% of band grades. The remainder of student grades comes from a combination of technique and other practices (mostly quizzes and practice logs). Within these criteria weights, grading appears to be rather generous. Band directors report giving "A's" to 75.4% of their students and "B's" to another 16.3%.

Question three: To What Degree Are the High School Band Assessment and Grading Practices Viewed as Adequate and Appropriate From the Band Director's Perspective?

While 76% of band directors agreed that their assessment methods are good enough to ensure quality instruction, 89.5% reported that, given more temporal resources, they would undertake more student learning assessment. The mean percentage of band grades that directors reported should come from attendance was significantly less than the mean for the actual current practice reported mean (25.7% versus 20.0%). The current percentage of band grades from band music performance and technique was significantly lower than what band directors reported the percentage should be. However, this

significant difference between mean current practice and mean ideal grade weighting resulted from a minority of participants. Most band directors indicated no difference between current grading practice and ideal grading.

Question four: Which of the Following Factors May be Associated With Variations in Reported Assessment and Grading Practices and/or Perspectives Regarding Those Practices: Regional Factors (MENC Region, Urban Versus Rural), School Factors (School Size, Number of Students Per Band), Band Director Factors (Educational Background, Years Experience, Major Instrument)?

Only band size, MENC region, band director educational background, and experience were found to be significantly related to assessment, grading, or satisfaction variables. The specific findings were as follows: (a) Directors of small bands, as well as directors with more education and experience, reported more tape recorder use (nearly 70% versus roughly 50%) in assessment; (b) the duration of performance assessments by band directors in the north central MENC region is significantly longer (267 seconds per assessment) than assessment durations in the western, south western, and north western regions (90 seconds per assessment); (c) the weight of grading criteria varies significantly between large and small bands in the MENC eastern, north central, and western regions. Specifically, in the western region, small band directors place less weight on participation; in the north central region, small bands put less weight on band music performance, and in the east region, directors of small bands put more weight on paper-and-pencil assessment.

It is possible that differences between MENC regions are a result of different cultural, political, or economic conditions specific to certain areas. For example, anecdotal evidence suggests that more public high school band programs in the upper mid west include free private lessons for all students than in other regions of the United States. This regional difference is likely the reason why the average assessment time in the north central region was almost three times longer than in the west, south west, or north west (teachers likely use private lesson time for individual band assessment).

#### Summary of Findings Regarding Best Practice

Having determined what assessment and grading strategies are being used by band directors and having determined that current practice is seen to meet local assessment needs in most cases, the relationship between current practice and best practice as described in the professional literature is discussed below (Goolsby, 1999; Killian, 1998; MENC, 1996).

#### Equipment Used for Assessment

While unassisted listening to student performance is a valuable form of assessment, reliability is greatly enhanced with the use of computer imaging of sound, audio and/or video recording (which allows for later comparison), and rubrics (Cope, 1996; Killian, 1998).

#### Computer-assisted Assessment

Although many band directors report using a computer to organize grades (55%), and several more report use of notation software, sequencing software, and theory tutoring and testing software, none of the participants reported using sound identification

and visualization software. It is possible that the cost did not appear justified for the teaching situations (when students are completely lost, directors do not need computers to tell them that student's played the wrong note).

#### Audio/Video Tape Recording Use for Assessment

Despite many articles about the benefits and possible uses of audio and video recording (Carlin, 1996; Goolsby, 1999; Killian, 1998; Robinson, 1995; Rutkowski, 1994), only 60% of the participant directors reported using audio tape for assessment (28% use video). Even fewer report using this technology for individual student assessment (33% audio, 7% video). A possible reason why two out of three band directors do not record individual student performance is time. Goolsby, (1999) suggested listening to student tapes two hours per day (including weekends) in order to assess 20 minutes of student performance per month. In the present study, 9.8% of band directors reported listening to each student for 20 minutes per month. None reported doing so with the use of a tape recorder.

#### Use of Performance Based Assessment

A small number of band directors reported using performance based assessment, several of whom included copies with returned surveys. In most cases, the rubrics appeared to be developed and used across single school districts. Most band directors did not report use of rubrics in performance based assessment. It is likely that band directors did not feel prepared to generate their own rubrics and simply did not use rubrics if none were provided. This may be an area that could be examined by university schools of music education.

### Comparison with Past Studies

Although the purpose of the study was not to examine long-term assessment changes, it is interesting to compare current grading practices with those reported almost a decade ago. McCoy (1991) divided grading into four divisions: cognitive, psychomotor, affective, and non-music. Each of these divisions was reached by summing a number of detailed grading criterion. The individual criterion for the non-music and affective divisions fall roughly into what would be called attendance and participation in the present study. The psychomotor criterion would generally fall into band music performance and technique in the present study (performance based assessment). The cognitive criterion could be placed in the Other (primarily paper-and-pencil) section of the present study. Figure 6 shows a comparison between the 1991 McCoy results and the results of the present study. Although the McCoy study was limited to a single state and there are many differences between the two studies, it appears that assessment among band directors has not change dramatically in the last decade.

### Implications

Current findings indicate that a small number of band directors are working hard to assess student learning in creative ways. The use of journals, video portfolios, self-critique papers, varied uses of tape recorder, and many other creative assessment strategies illustrate the work that some band directors are doing to improve their assessment. Many creative grading ideas such as peer grading and community service

requirements point to the effort a few dedicated band directors are putting into their grading systems. Unfortunately, as illustrated in Figure 6, band programs

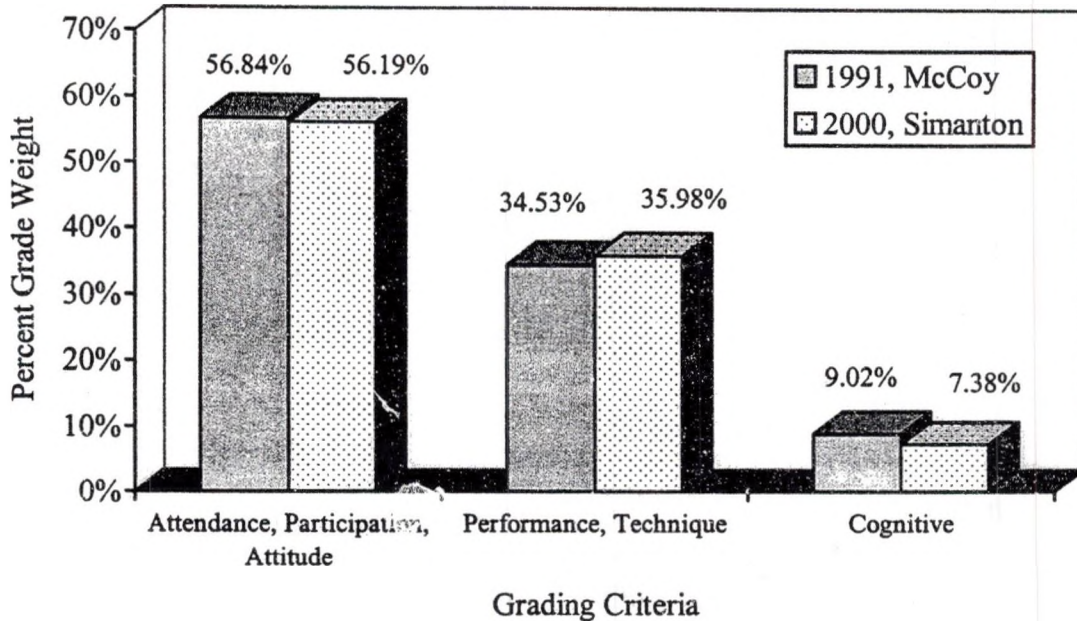


Figure 6. Comparisons between the McCoy (1991) study and the present study on the basis of grade weightings.

as a whole have not improved in the last decade with regard to grading. Deficiencies in assessment and grading described by Lehman in 1992 are still present today.

The workload of band directors appears to be immense. Band teachers reported directing an average of 2.73 bands besides teaching up to 14 other ensembles and classes besides band ( $M = 2.41$ ). Many band directors reported working alone with multiple bands of 100 students or more. Many of these band directors have little opportunity for assessment other than group assessment (Colwell, 1991). Directors of small bands reported audio taping student performance significantly more often than directors of large

bands. Band directors with more education and experience were also more likely to tape record their students. This could be a result of more educated and experienced band directors placing more value on audio-taping students or may simply be a result of less experienced band directors feeling overwhelmed by other work and not finding the time to assess students as their more experienced peers do. Almost 9 out of 10 band directors indicated that they would do more assessment if they had the time (89.6%).

Many band directors appear to be satisfied with current assessment and grading practice. More than three quarters of the study participants agreed that their current assessment was good enough to meet the current educational needs (76.0%). Although the mean grade weighting variables show significant differences between current practice and best practice (significantly less grade weight on attendance and more on performance and technique), a closer look at the results shows that over half of the respondents' current practice grade weighting and ideal grade weighting for all criteria were the same.

Despite the emphasis placed on assessment and accountability during the last ten years, the publication of National Standards for Arts Education (Consortium of National Arts Education Associations, 1994), and the many efforts made to improve instrumental music assessment, no indications are available to demonstrate significant changes in the way assessment and grading take place in high school band programs (McClung, 1997; McCoy, 1991).

A possible reason why band directors apparently have not changed (and may not be motivated for future change) regarding assessment and grading is the apparent local satisfaction with current practice. Less than 1% of the participants disagreed that school

officials were satisfied with current assessment practices, and 2% reported that school officials were dissatisfied with current grading practices. While some the national accountability movement is being felt at the local level in subjects such as math and reading, band directors do not seem to have yet come under scrutiny at the local level. Furthermore, it appears that if there is no local impetus for change, it is unlikely that change will take place.

Band directors may also be simply waiting to see what happens before putting forth the effort to change. Over half of the study participants reported that their assessment and grading practices were similar to those of other band directors they knew and another third reported that they did not know. Only 13% thought that their assessment and grading practices were different from most other band directors.

One of the important findings of the study was the difference between band directors with master's degrees and those with only bachelor's degrees. Although only use of tape recorder was statistically significant, trends were evident in many of the assessment and grading variables. A possible explanation is that discussion of assessment and grading may be a more important part of graduate programs than undergraduate programs. This finding suggests that changed emphasis in higher education may initiate changes at the high school level.

#### Recommendations for Further Research

On the basis of this study, the following investigations are suggested:

1. An in-depth study of the assessment processes (formal and informal) used for performance assessment in large and small high school bands.



2. An investigation of the utility (time efficiency and accuracy) of a variety of performance assessment strategies used in high school bands.

3. An investigation into the relationship between educational background and grading and assessment to determine if any specific graduate or undergraduate courses and/or content may be related to the use of specific assessment and grading practices.

4. An investigation of regional differences in high school band programs, possible social, cultural, or political explanations, and possible application to other regions.

5. An investigation of the differences between large and small bands including differences in assessment and grading policies and possible explanations of those differences.

6. An investigation of assessment and grading practices from the perspectives of students, parents, and principals including regional differences.

7. A comparison of the real educational and musical impact of best practice assessment and grading.

8. A longitudinal study of assessment and grading practices to determine if, in fact, practices are changing nationally and/or regionally.

**APPENDIX A**  
**SURVEY INSTRUMENT**

## Band Assessment and Grading Survey

### About You

- (1) \_\_\_\_ Years experience (teaching band)  
 (2) \_\_\_\_ Years in your current position

Check all that apply regarding your educational background:

- (3) Bachelors  
 (4)        in music  
 (5)        in music education  
 (6) Masters  
 (7)        in music  
 (8)        in music education  
 (9) Doctorate  
 (10)       in music  
 (11)       in music education  
 (12) \_\_\_\_\_ Major Instrument

### City/School Band Information

- (13) \_\_\_\_\_ Population of city/town  
 (14) \_\_\_\_ Number of high schools in city/town  
 (15) \_\_\_\_ Number of students in high school

Check the socio-economic status of most students in your school:

- (16) Lower    Middle    Upper  
 (17) \_\_\_\_ Number of bands in your school  
 (18) \_\_\_\_ Number of other instrumental ensembles in your school (Jazz bands, quintets, etc.)  
 (19) \_\_\_\_ % Percentage high school students who participate in the band program  
 (20) \_\_\_\_ Number of band teachers (counting part time as .5)

- (21) \_\_\_\_ Number of bands you teach (or co-teach)

- (22) \_\_\_\_ Number of other classes (besides band) you teach

*For the rest of the survey we will look in-depth at assessment and grading as it takes place in just one band. If your school has several, please provide information as applies only to the band that contains the most 11<sup>th</sup> grade students*

- (23) \_\_\_\_ Number of students in band  
 (24) \_\_\_\_ Number of years instrumental experience among most students in this band  
 (25) \_\_\_\_ Number of students who play in other ensembles (Jazz ensemble, quintets, etc.)

Other grade levels (besides 11<sup>th</sup>) present in band

- (26) 9<sup>th</sup>  
 (27) 10<sup>th</sup>  
 (28) 12<sup>th</sup>  
 (29) Other \_\_\_\_\_

### Assessment Information

#### Frequency and type of assessment

- (30) \_\_\_\_ How many times per month is each student's performance graded within rehearsals? (other than attendance or participation)  
 (31) \_\_\_\_ How many times per month is each student's performance graded outside of rehearsals?  
 (32) \_\_\_\_ When you are listening to individual students perform, how many seconds (on average) do you spend listening to each student?

**Indicate equipment used for assessment**

- (33) Tape Recorder  
 (34) to record rehearsal  
 (35) \_\_\_\_\_ times/month  
 (36) to record individuals  
 (37) \_\_\_\_\_ times/ month
- (38) Video camera  
 (39) to record rehearsal  
 (40) \_\_\_\_\_ times/month  
 (41) to record individuals  
 (42) \_\_\_\_\_ times/ month
- (43) Computer  
 (44) to organize grading  
 (45) to analyze student performances  
 (46) Using what software?  
 \_\_\_\_\_
- (47) Paper/pencil  
 (48) for quizzes  
 (49) for journaling  
 (50) Other (Specify)  
 \_\_\_\_\_
- (51) Other equipment (Explain)  
 \_\_\_\_\_  
 \_\_\_\_\_

**Grading Information**

Who is informed of band grading policies and how? (Check all that apply)

- (52) Students  
 (53) verbally  
 (54) in writing
- (55) Parents  
 (56) verbally  
 (57) in writing
- (58) Principal  
 (59) verbally  
 (60) in writing

What percentage of band grades currently comes from each of the following:

- (61) \_\_\_\_\_ % Attendance  
 (62) \_\_\_\_\_ % Participation/Attitude  
 (63) \_\_\_\_\_ % Performance of band music  
 (64) \_\_\_\_\_ % Technique and or sight- reading  
 (65) \_\_\_\_\_ % Other (Explain) \_\_\_\_\_  
 100 % Total

Estimate the percentage of students who receive the following grades in band in a typical grading period:

- (66) \_\_\_\_\_ % A  
 (67) \_\_\_\_\_ % B  
 (68) \_\_\_\_\_ % C  
 (69) \_\_\_\_\_ % D or F

**Satisfaction with Current Practice**

What percentage of band grades should come from each of the following:

- (70) \_\_\_\_\_ % Attendance  
 (71) \_\_\_\_\_ % Participation/Attitude  
 (72) \_\_\_\_\_ % Performance of band music  
 (73) \_\_\_\_\_ % Technique and/or sight-reading  
 (74) \_\_\_\_\_ % Other (Explain) \_\_\_\_\_  
 100 % Total

*Please rate the following statements using the following scale:*

SA = strongly agree, A = agree, N = neutral, D = disagree, SD = strongly disagree

- |   |             |
|---|-------------|
| (75) My assessment methods are good enough to ensure quality instruction                              | SA A N D SD |
| (76) I am concerned primarily with ensemble rather than individual assessment                         | SA A N D SD |
| (77) Parents in my school seem to be satisfied with current band <u>assessment</u> practices          | SA A N D SD |
| (78) Band students in my school seem to be satisfied with current band <u>assessment</u> practices    | SA A N D SD |
| (79) School officials in my school seem to be satisfied with current band <u>assessment</u> practices | SA A N D SD |
| (80) Parents in my school seem to be satisfied with current band <u>grading</u> practices             | SA A N D SD |
| (81) Band students in my school seem to be satisfied with current band <u>grading</u> practices       | SA A N D SD |
| (82) School officials in my school seem to be satisfied with current band <u>grading</u> practices    | SA A N D SD |
| (83) I would do more student learning assessment if I had more time                                   | SA A N D SD |
| (84) I think it is <u>unfair</u> to grade students by how well they play                              | SA A N D SD |
| (85) My assessment and grading practices are similar to those of most of the band directors I know    | SA A N D SD |

**APPENDIX B**

**COVER LETTER FOR FIRST MAILING**

Dear Colleagues,

Assessment and grading have recently become hot topics in music education and education in general. For example, the entire September 1999 Music Educators Journal was focused on assessment in music education. However, little is known about attitudes and assessment practice among working music education professionals. Based on recent conversations with band directors I have found that band directors have widely varying views regarding assessment and grading practices. Some believe that assessment and grading are over-emphasized in the current political/educational climate. Others see assessment and grading as key elements that will help ensure a place for music in education. The purpose of this study is to measure band directors' practices and attitudes regarding assessment and grading both regionally and nationally.

You are one of 600 high school band directors randomly selected from all across the United States to be part of this study of high school band assessment and grading. In order to make accurate comparisons between the 6 MENC regions, the cooperation and participation of all participants are very important.

Please take a few minutes to complete and return the enclosed survey. If you are not certain regarding exact answers to some items (city population, school enrollment, etc.), please provide your best estimate. All responses will be confidential and anonymous. Surveys have been marked to allow a second mailing. All identifying information will be destroyed at the end of the data collection phase of the study. Completion and return of the survey represents consent to participate in the study. Not enough is known about practice in the field to make definitive judgments about one set of practices or another. I am only interested in what you are doing now and how satisfied you are with what you are doing. No judgments about one practice or another will be undertaken. Your participation will greatly help in answering many questions regarding assessment practices regionally and nationally.

Thank you for your participation.

Sincerely,

Edward G. Simanton  
Ph.D. Candidate  
University of North Dakota

**APPENDIX C**

**COVER LETTER FOR SECOND MAILING**



Dear Colleagues,

This is a follow-up to the survey you should have received about a month ago. If you have already completed and mailed that survey, ignore this note. If you have not yet mailed in your survey, please take a moment to complete the enclosed survey; it is extremely important to my dissertation research.

Assessment and grading have recently become hot topics in music education and education in general. For example, the entire September 1999 Music Educators Journal was focused on assessment in music education. However, little is known about attitudes and assessment practice among working music education professionals. Based on recent conversations with band directors I have found that band directors have widely varying views regarding assessment and grading practices. Some believe that assessment and grading are over-emphasized in the current political/educational climate. Others see assessment and grading as key elements that will help ensure a place for music in education. The purpose of this study is to measure band directors' practices and attitudes regarding assessment and grading both regionally and nationally.

You are one of 600 high school band directors randomly selected from all across the United States to be part of this study of high school band assessment and grading. In order to make accurate comparisons between the six MENC regions, the cooperation and participation of all participants is very important.

Please take a few minutes to complete and return the enclosed survey. If you are not certain regarding exact answers to some items (city population, school enrollment, etc.), please provide your best estimate. All responses will be confidential and anonymous. All identifying information will be destroyed at the end of the data collection phase of the study. Completion and return of the survey represents consent to participate in the study. Not enough is known about practice in the field to make definitive judgments about one set of practices or another. I am only interested in what you are doing now and how satisfied you are with what you are doing. No judgments about one practice or another will be undertaken. Your participation will greatly help in answering many questions regarding assessment practices regionally and nationally.

If you prefer to respond online, this survey is found at the following website:

[www.members.tripod.com/Simanton/music\\_survey.html](http://www.members.tripod.com/Simanton/music_survey.html)

At the beginning of the survey you will be asked for a survey code. Please type in the name of the state where you teach. This will identify you as a study participant and make it possible to aggregate your responses with that of others from your MENC region. If you have trouble finding the site, be sure that you are using a capital "S" in Simanton (it won't work with a small "s").

Thank you for your participation.

Sincerely,

Edward G. Simanton  
Ph.D. Candidate  
University of North Dakota

## REFERENCES

American Council for the Arts. (1995). Measuring up to the challenge: What standards and assessment can do for the arts. New York: Author

Anderman, E. & Maehr, M. (1994). Motivation and schooling in the middle grades. Review of Education Research, 64(2), 287-309.

Birge, E. (1928). History of Public School Music in The United States of America. Boston: Oliver Ditson.

Boyd, W. & Kerchner, C. (Eds.). (1988). The Politics of excellence and choice in education: 1987 yearbook of the politics of education association. Philadelphia: Falmer.

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

Britton, A. (1991). American music education: Is it better than we think? A discussion of the roles of performance and repertory, together with brief mention of certain other problems. In Richard Colwell, (Ed.) Basic concepts in music education II (pp. 175-190). Niwot, CO: University Press of Colorado.

Carlin, J. (1996). Videotape as an Assessment Tool. Teaching Music, 3(4), 38-39, 54.

Casey, D. (1992). Descriptive research: Techniques and procedures. In Richard Colwell (Ed.), Handbook on research on music teaching and learning (pp. 115-123). New York: Schirmer.

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

Colwell, R. (1970a). The evaluation of music teaching. Englewoods Cliffs, NJ: Prentice-Hall.

Colwell, R. (1970b). Music achievement tests 3 and 4. Chicago: Follett Educational.

Colwell, R. (1991). Evaluation. In Richard Colwell, (Ed.) Basic concepts in music education II (pp. 246-278). Niwot, CO: University Press of Colorado.

Consortium of National Arts Education Associations. (1994). National standards for arts education. Reston, VA: MENC.

Crooks, T. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58 (4), 438-481.

Cope, C. (1996). Steps toward effective assessment. Music Educators Journal, 91 (11), 39-42.

Davidson, J. (1993). Visual perception of performance manner in the movements of solo musicians. Psychology of Music, 21, 103-113.

Farnum, S. (1969). Farnum string scale. Winona, MN: Hall Leonard.

Flynn, P. (1995). Global Competition and Education: Another Sputnik? Social Studies, 86(2) 53-55.

Freedman, M. (1965). A technique for the analysis of musical instrument tones. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Freedman, D., Pisani, R., & Purves, R. (1980). Statistics. New York: W. W. Norton & Company, Ltd.

Fullan, M. (1993). Innovation, reform and restructuring strategies. in Cawelti, G. (Ed.) Challenges and achievements of American education. Alexandria, VA: Association for Supervision and Curriculum Development.

Goolsby, T. (1999). Assessment in instrumental music. Music Educators Journal, 86(2) 31-35, 50.

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

Gordon, E. (1991). Iowa tests of music literacy: Revised. Chicago: G.I.A.

Grieshaber, K. (1993). Development of a comprehensive computerized scoring method for measuring rhythmic performance. Bulletin of the Council for Research in Music Education, 119, 127-136.

Graham, S. L. (1989). A comparison of group achievement on the Manitoba music assessment program and variables related to school music programs in Manitoba. (Masters thesis, Brandon University, 1989). Masters Abstracts International, 28-02, 0194

Gronlund, N. (1981). Measurement and evaluation in teaching. New York, NY: Macmillan .

Hoffer, C. Introduction to music education. Belmont, CA: Wadsworth, Inc.

Humphreys, J. (1989). An overview of American public school bands and orchestras before World War II. Bulletin of the Council for Research in Music Education, 101, 50-60.

Killian, J. (1998). Assessment in music performing groups: some practical suggestions. [Online]. Available: <http://arts.usf.edu/music/smef/art-jk.html>.

Lehman, P. (1992). Assessing learning in the music classroom. Bulletin of the National Association of Secondary School Principals, 76(544), 56-62.

Matheny, J. (1994). A grading system for secondary music groups. Music Educators Journal, 81(1), 37-39.

McClung, A. (1997). A descriptive study of learning assessment and grading practices in the high school choral music performance classroom. Dissertation Abstracts International, 57/08, 3437A.

McConathy, O., Gehrkens, K., & Birge, E. (1921). Present status of music instruction in colleges and high schools (United States Bureau of Education, Bulletin no. 9). Washington: U.S. Bureau of Education.

McCoy, C. (1988). An exploratory study of grading criteria among selected Ohio ensemble directors. Contributions to Music Education, 15, 15-19.

McCoy, C. (1991). Grading students in performing groups: a comparison of principal's recommendations with director's practices. Journal of Research in Music Education, 39(3), 181-190.

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

McPherson, G. (1994). Factors and abilities influencing sight-reading skill in music. Journal of Research in Music Education, 42(3) 217-231.

McPherson, G. (1995). The assessment of musical performance: Development and validation of five new measures. Psychology of Music, 23(2) 142-161.

MENC. (1994). Perspectives on implementation: Arts education standards for America's students. B. O. Boston (Ed.). Reston, VA: Author

MENC. (1996). Performance standards for music: Strategies and benchmarks for assessing progress toward the national standards. Reston, VA: Author.

Monroe, R. (1995). Music education in Ohio high schools: a comparative study of the opinions of music teachers, principals, and professors of music education. (Doctoral dissertation, Miami University, 1994) Dissertation Abstracts International, 56-01A, 16760.

Montgomery, S. (1994). Minds for the making. New York: Guilford.

National Center for Education Statistics. (1998). NAEP and music: Framework, field test, and assessment. Washington, DC: Author.

National Education Association. (1996). Estimates of school statistics as provided by the State Departments of Education, 1995-96. Washington: Author

Nitko, A. (1996). Educational assessment of students. Englewood Cliffs, NJ: Prentice-Hall.

Peters, G. (1974). Feasibility of computer-assisted instruction for instrumental music. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign).

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

Phelps, R., Ferrara, L., & Goolsby, T. (1993). A guide to research in music education (4th ed.). Metuchen, NJ: Scarecrow Press.

Popham, W. (1973). Evaluating instruction. Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. (1993). Educational evaluation (3rd ed.). Needham heights, MA: Allyn and Bacon

Pulliam, J. (1991). History of education in America (5th ed.). New York, NY: Merill.

Radocy, R. (1998). Review of the Iowa Tests of Music Literacy, Revised. In J. Impara & B. Plake (Eds.), The Thirteenth Mental Measurements Yearbook. (pp. 551-555) Lincoln, NE: University of Nebraska Press.

Rees, F. & Michelis, R. (1991). Identifying musical performance behavior in instrumentalists using computer-based sound spectrum analysis. Bulletin of the Council for Research in Music Education, 110, 53-66.

Robinson, M. (1995). Alternative assessment techniques for teachers. Music Educators Journal, 81(5) 28-34.

- Russo, M. (1988). Written evaluations. Music Educators Journal, 75(2), 38-39.
- Rutkowski, J. (1994). Assessment in music performance classes. Music Educators Journal, 80(3), 53-54.
- Sarason, S. (1995). School change: the personal development of a point of view. New York: Teachers College Press.
- Schleuter, S. (1984). A sound approach to teaching instrumentalists. Kent, OH: Kent State University Press.
- Shepard, L. (1989). Why we need better assessments. Educational Leadership, 46(7), 4-9.
- SPSS Inc. (1999). SPSS for windows version 10.0.5 [Computer program] SPSS Inc: Chicago, IL.
- Swanwick, K. (1994). Music knowledge: Intuition, analysis, and music education. New York: Routledge.
- Watkins, J. & Farnum, S. (1954). The Watkins-Farnum performance scale: Form a. Winona, MN: Hall Leonard.
- Watkins, J. & Farnum, S. (1962). The Watkins-Farnum performance scale: Form b. Winona, MN: Hall Leonard.
- What's New in the Music Industry. (1997). Instrumentalist, 51(7), 38-44.
- Wolverton, V. (1992). Comprehensive musicianship revisited: Research-based applications for the choral music classroom. Update, 10(2), 10-15.
- Zdzinski, S. (1991). Measurement of solo instrumental music performance: A review of literature. Bulletin of the Council for Research in Music Education, 109, 47-58.
- Zdzinski, S. (1996). Parental involvement, selected student attributes, and learning outcomes in instrumental music. Journal of Research in Music Education, 44(1), 34-48.