Sight-reading for piano students: Comparing three methods of assessment
Lemay, Catherine
ProQuest Dissertations and Theses; 2008; ProQuest Dissertations & Theses (PQDT)
pg. n/a

uOttawa

L'Université canadienne
Canada's university

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Catherine Lemay
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Music)
GRADE / DEGREE

Department of Music
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Sight-reading for Piano Students: Comparing Three Methods of Assessment

TITRE DE LA THÈSE / TITLE OF THESIS

Gilles Comeau
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Elaine Keillon

Ramesh Balasubramaniam

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sight-reading for Piano Students: Comparing Three Methods of Assessment

Catherine Lemay

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillments of the requirements
for the MA Degree in Music

Faculty of Music
University of Ottawa
September 2008

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-48612-2
Our file Notre référence
ISBN: 978-0-494-48612-2

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

© Catherine Lemay, Ottawa, Canada, 2008

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sight-reading for Piano Students: Comparing Three Methods of Assessment

Table of Contents

Table of Contents....ii
List of Tables....v
List of Figures....vi
Abstract....vii

1. Introduction....1
   1.1. Context....2
      1.1.1. Defining Music Reading and Sight-reading....2
      1.1.2. Importance of Music Reading....6
      1.1.3. Difficulty of Music Reading....10
      1.1.4. Need for Further Research....11
   1.2. Review of Literature....12
      1.2.1. Measurement by Tests and Scales....13
         1.2.1.1. Belwin-Mills Singing Achievement Test....13
         1.2.1.2. Watkins-Farnum Performance Scale....14
      1.2.2. Sight-Reading Error Quantifications....20
         1.2.2.1. Eaton's Grading Instructions....20
         1.2.2.2. Gilman's Scoring Algorithm....22
         1.2.2.3. Gudmunsdottir's Error Classification....24
         1.2.2.4. Salis' Error Categorization....27
      1.2.3. Evaluation by Expert Examiners....29
   1.3. Purpose of the Study....31

2. Methodology....33
   2.1. Developing a Scale....35
      2.1.1. Finding Musical Stimuli....35
      2.1.2. Revising the Grading System....40
      2.1.3. Adapting the Error Characterization....43
   2.2. Comparing Assessment Methods....51
      2.2.1. Participants....51
      2.2.2. Procedure....52

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3. Data Presentation of Performance Assessment.....55
3.1. Measurement According to the Piano Sight-reading Scale.....58
3.1.1. Measurement of Subjects.....60
3.1.2. Subject Placement.....63
3.2. Quantification According to Gilman's Scoring Algorithm.....66
3.2.1. Quantification of Subjects.....68
3.2.2. Subject Placement.....74
3.3. Evaluation by Expert Examiners.....75
3.3.1. Subject Placement.....77
3.4. Conclusion.....80

4. Data Comparison.....81
4.1. Comparison of Assessment Procedures.....81
4.1.1. Types of Errors.....81
4.1.2. Allocation of Marks.....82
4.1.3. Length and Complexity.....84
4.1.4. Conclusion.....85
4.2. Comparison of Subject Assessment.....85
4.2.1. Grade Deductions.....86
4.2.2. Final Grades.....90
4.2.3. Subject Placement.....94
4.2.4. Conclusion.....97

5. Discussion.....99
5.1. Comparison of Assessment Procedures.....99
5.1.1. Types of Errors.....99
5.1.2. Allocation of Marks.....101
5.1.3. Length and Complexity.....102
5.2. Comparison of Subject Assessment.....104
5.2.1. Grade Deductions.....104
5.2.2. Final Grades.....109
5.2.3. Subject Placement.....110

6. Conclusion.....113

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

References.....**118**
Appendixes.....**125**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Tables[1]

1. Sight-reading Assessment Methods.....1
2. Review of Literature Organization.....12
3. Eaton's Grading Instructions.....22
4. Gilman's Scoring Algorithm.....24
5. Gudmunsdottir's Error Classification 26
6. Salis' Error Categories.....28
7. Requirements for Sight-reading as Listed in the 2001 Edition of the Royal Conservatory of Music Piano Syllabus.....37
8. Grading Measure for the Piano Sight-reading Scale.....42
9. Watkins-Farnum Performance Scale.....45
10. Piano Sight-reading Scale Error Classification.....49
11. Participants Characteristics.....51
12. CuBase Specifications for Each Exercise.....54
13. Measurement of Subjects with the Piano Sight-reading Scale.....59
14. Subject Placement with the Piano Sight-reading Scale.....64
15. Average Subject Placement with the Piano Sight-reading Scale.....65
16. Quantification of Subjects with Gilman's Scoring Algorithm.....67
17. Subject Placement with Gilman's Scoring Algorithm.....74
18. Evaluation of Subjects by Expert Examiners.....76
19. Subject Placement with Expert Examiners.....78
20. Average Subject Placement with Expert Examiners.....79
21. Grading Units According to Each Method of Assessment.....91
22. Final Grades Obtained with Each Method of Assessment.....92
23. Ranking with the Three Methods of Assessment.....95
24. Placement as Good or Poor Reader with Each Method of Assessment.....97

[1] Tables 3, 4, 5, 6, 7, and 9 have been built by this author to present, in a concise way, grading procedures described by each method.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Figures

1. Watkins-Farnum Performance Scale: Average Sores.....16
2. Watkins-Farnum Performance Scale: Grading Chart 17

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Abstract

One important component of music reading research is measurement, quantification, and evaluation of accuracy in sight-reading performance. Researchers have used various methods of assessment such as the sight-reading tests and scales, quantification of errors, and evaluation by expert examiners. These three methods of assessment have been used independently in research; however, they have never been tested to determine if they provide comparable analyses. This study, therefore, adapted the Watking-Farnum Performance Scale for wind instrumentalists to be used in the context of piano performance and then compared it to the two other assessment methods, more specifically Gilman's Scoring Algorithm and Expert Examiners. Each method of assessment was used to analyze the sight-reading performances of eight piano students on five newly composed exercises. The assessment procedures and subject assessments were then compared. It was found that these methods differed greatly in their assessment procedures as well as in their assessment of subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1

1. Introduction

In music education, one key skill when learning a musical instrument is the ability to read musical notation. Sight-reading is an important means for practicing as well as assessing music reading ability. The purpose of this thesis is to provide an in depth comparison of measurement, quantification, and evaluation methods commonly used for the assessment of sight-reading performance by researchers studying music reading. The terms measurement, quantification, and evaluation have very specific meanings in the context of this research. Measurement is used for the methodical assignment of numbers to represent quantitative characteristics of performance, specifically for assessment with sight-reading scales and tests. Quantification is used synonymously to measurement, however, will be reserved for assessment with sight-reading error quantification methods. Evaluation is used in the sense of judging performance with a set of subjective criteria, more specifically, as done in performance evaluation by expert examiners or teachers. Assessment will be used in a general sense encompassing the estimation, judging, measurement, and evaluation of performance (Bloch et al., 2002).

Table 1. Sight-reading Assessment Methods

Sight-reading Assessment Methods	Measurement by sight-reading tests and scales
	Sight-reading error quantifications
	Evaluation by expert examiners

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The following chapter will first place music reading in context, by defining both music reading and sight-reading, by presenting the importance and difficulty of mastering this skill, and then highlighting the need for further research. It will then provide a review of literature of the various methods which have been used by researchers to assess sight-reading performance, organized in the following three categories: measurement by tests and scales; sight-reading error quantifications; and evaluation by expert examiners. This chapter will conclude by presenting the purpose of the present study.

1.1. Context

To contextualize music reading, we will first define music reading and sight-reading, we will then demonstrate both the importance of this skill and the difficulty of mastering it. We will outline the need for sight-reading assessment tools and review the measurement, quantification, and evaluation methods that have been used in music reading research.

1.1.1. Defining Music Reading and Sight-reading

The skill of music reading is difficult to characterize but we will present a few definitions by scholars, musicians, and educators. In 1960, Petzold defined music reading as "the process of reading and interpreting the various kinds of music symbols and converting these symbols into sound" (p. 271). Serafine (1988) characterizes the skill as the "act of decoding notation and coordinating muscular action based on thought processes that are nonauditory" (p. 72). Hodges (1992) defines music reading as "the process of converting special visual symbols - music notation - into sounds" (p. 466). Scripp (1995) proposed that "skill in music reading ... requires both the ability to perform efficiently and effectively from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

notation as well as the ability to detect errors in performance from the printed score" (p. 1). Stewart, *et al.* (2003) suggests that "music reading for keyboard performance can [...] be envisaged as a visuospatial sensorimotor task, in which stimuli that vary along a vertical dimension are mapped onto the fingers, which strike horizontally arranged response elements" (p. 71). Finally, Ely and Rashkin (2005), in the *Dictionary of Music Education*, define music reading as "the process of seeing music notation, interpreting the notation, and performing accordingly", therefore, "music reading is the process of identifying visual cues (notes, expression markings, tempo indications, etc.) and performing a series of physical and mental actions necessary to produce the desired sound or sounds" (p. 284).

Many of the music reading definitions introduced above provide interesting points on this skill, such as the interpretation of musical symbols and their conversion into sound, that it is a task requiring both ocular involvement and motor action, the need to detect errors, among other. However, many do not provide a complete explanation of this task, lacking one or more facets which we feel are necessary. Elements such as the interpretation of musical notation, the physical component, as well as the auditory component of this skill are among the few facets missing from some of the definitions. Consequently, for the purpose of this research, Ely and Rashkin's (2005) definition will be retained because of its accuracy and clarity. It includes all the elements necessary to the task in question, perceptual, cognitive, kinesthetic, and auditory, and will meet the requirement of this study. This definition serves as the starting point from which to initiate the research which focuses on the assessment of music reading.

The task of sight-reading is clearly linked to the reading of musical notation and has, therefore, served as a means of assessing the music reading process. Consequently, it is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

necessary to clearly define sight-reading as it is the skill being evaluated in this study. Elkan (1948), in her guide to piano sight-reading, provides the following definition of sight-reading: "The skill and ability to play accurately as you read a printed page of piano music, whether you see it for the first time (prima vista) or whether you have read and played it many times before" (p. 5). Wolf (1976) states that "music sight-reading [is] the ability to read from a score without benefit of practice" (p. 143). He then adds that this task is "a complex process which appears to involve at least two distinct skills: the first is the reading skill—the musician must scan and process the printed music and the second is a mechanical skill—he must place his fingers in the right place on his instrument at precisely the correct moment" (p. 143). Gabrielsson (2003) also believes that "sight-reading means performing from a score without any preceding practice on the instrument of that score, to perform *a prima vista*" (p. 243). Lehmann and McArthur (2002), describe sight-reading as follows: "The ability to perform with little or no rehearsal ... one could limit the description of sight-reading somewhat by requiring that the music be physically played (gestured, softly voiced, or otherwise sounded) at an acceptable tempo and with appropriate expression, thereby excluding the mere deciphering of the notation, especially the tediously slow groping for notes with nothing more in mind than to internalize the piece ... one could restrict sight-reading to those activities where the performer intends to approximate the final product as closely as possible" (pp. 135-136). Thompson and Lehmann (2004) define sight-reading as "the practice of playing a piece of music directly from the score on first encounter or after brief rehearsal" (p. 145). Ely and Rashkin (2005) suggest that sight-reading is "the ability to accurately perform or interpret music notation that has not been seen before" (p. 409). Finally, Stauffer (2005) characterizes the reading of music at first sight by:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> The ability to translate musical symbols and terms into a musical performance using strict time limits. The musician must quickly perceive symbols ahead while playing the symbols that he or she has just perceived. This seems to imply that the performer must be thinking about several things at the same time, but it is impossible to concentrate on more than one thing at any instant. Perceptions must trigger automatic or semiautomatic physical responses, so the mind can perceive the next musical symbols. All of this needs to be accomplished with a continuing sense of rhythmic movement in time. (p. 21)

For the purpose of this study, our position in relation to sight-reading is that it is the first reading of a piece, for that reason Elkan's description, which does not fit our interpretation of this skill, will not be kept. Though Lehmann and McArthur's definition also suggested that sight-reading could involve some rehearsal it was not entirely dismissed as it contained other elements of interest. Many of the other definitions, though they partially covered our characterization of sight-reading, were not sufficiently thorough some leaving out such elements as interpretation of musical symbols, the sense of timing, and the intent to reproduce the final product as closely as possible. Therefore, to guide our research we will adopt Stauffer's definition of sight-reading in part, and combine it with a portion of Lehmann and McArthur's. From Stauffer's definition we will retain the need to translate musical symbols into a musical performance while respecting a continued sense of rhythmic movement in time. The element borrowed from Lehmann and McArther's will be the performer's intent to generate a performance which will be as close as possible to the final and perfected performance of a musical work. These definitions together will formulate a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

well-rounded depiction of the sight-reading skill which unites all elements of the cognitive process from the deciphering of musical notation to the final product.

In this study, the task of sight-reading will be used for assessment purposes and for that reason it was necessary to provide a clear definition. However, in most research involving the assessment of these skills, music reading and sight-reading are often discussed interchangeably. Nevertheless, in this project music reading will be used to define the act of decoding and interpreting musical notation while performing according to all visual cues and sight-reading will be used specifically to define a task which requires the music reading of new notation within a timing constraint as presented in the previous paragraph.

1.1.2. Importance of Music Reading

In the Western classical music tradition, knowing how to read musical notation is essential. Accordingly, it is seen as a fundamental competence to develop as part of any musical training (Agay, 2004; Galyen, 2005; Schmoll, 1912; Sloboda, 1978, 1988, 2005). Ettore Mazzoleni (Berlin, 1986), a leading figure in the development of Canada's musical education, stated that "to be able to read at sight is of the first importance to every piano student" (p. 2). Rubinstein (1950), one of the twentieth century's leading piano performers, compared the importance of learning to sight-read music to the imperative need for the literary scholar to be able to read text with ease. A few years later, Christ (1953) reinforced the importance of music reading by stating that "music is a language and should be studied as a language, and both the eye and the ear must be cultivated. As in a language, we acquire facility by speaking and reading in that language, so in music we must have abundant material and use it abundantly" (p. 3). For many musicians and educators, it is therefore

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

essential for young pianists of the Western classical tradition to be musically literate in order to acquire full membership in its society (Sloboda, 1978). The importance of music reading has also been supported by many researchers and educators (Craige, 1993; Sloboda, 2005; Wiltshire, 2006).

The importance of music reading and sight-reading can be found everywhere in the culture of Western classical music education: in method books, textbooks, conservatory exams, music competitions, as well as in university music programs. Learning to read musical notation is one of the central elements of many piano methods for both young and late beginners (Jacobson, 2006). The teaching of music reading is widespread and at least three different methods have been developed: 1) the middle C approach (Berlin, 1984; Thompson, 2005), 2) the intervallic approach (Clark & Goss, 1993), and 3) the multiple key approach (Pace, 1954). More recently the eclectic method, which combines two or all three approaches to music reading, is becoming very popular (Faber, Faber, & McArthur, 2001). In addition, educational material have been produced with exclusive attention to the learning of music reading, such as *How to Read Music: Music Reading Made Simple* (Burrows, 1999), *Learn to Read Music* (Shanet, 1971), *The Music Road: A Journey in Music Reading* (Starr, 2000), *Music Reading for Keyboard* (Steelman, 1998), and *The Musician's Guide to Reading and Writing Music* (Stewart, 1999). An extensive number of sight-reading methods have also been published, including *Complete Series of Sight-reading and Ear Tests* (Bennett & Capp, 1969), *Four Star Sight Reading and Ear Tests: Daily Exercises for Piano Students* (Berlin, 1986), *Sight-Reading Made Easy: A Complete Graded Piano Course* (Bradley & Tobin, 1947), *The Magic Reader* (Guhl, 1989), *Right@Sight* (Johnson, 2001), and *Sight-Reading* (Kember, 2004). Learning to read music is of such importance that it is a key

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

module of many music education textbooks: Agay (2004), in the chapter "Essentials of Piano Performance" of his textbook, primarily discusses music reading components; Baker-Jordan (2004) provides extensive review of piano methods in which the approach to music reading occupies an significant place; Bastien (1995) dedicates over half of his book to a survey of beginner piano methods in which a logical and practical sight-reading program is listed as a fundamental feature; Camp (1992) devotes an entire chapter to the teaching of music reading; Gordon (2001) provides a section on the instruction of sight-reading; both Jacobson (2006) and Rabin (1995) grant an entire chapter of their music textbooks to the teaching and learning of rhythm and reading; and finally, Uzler, Gordon, and Mach (1991) dedicate an entire section of their book to the teaching of music reading.

Music reading is a skill that is not only at the forefront of early piano learning, it remains an important ability throughout a musician's training. Music schools and conservatories have deemed music reading an essential skill, making it part of their examination syllabus through the sight-reading component (Babin, 2005). Conservatories across Canada allocate approximately 10 percent of the total mark of the practical exam to sight-reading: the Canadian National Conservatory of Music assigns 12 marks out of 100; both the Royal Conservatory of Music of Toronto and the Victoria Conservatory of Music allocate 10 marks out of 100; Conservatory Canada gives between 8 and 10 marks out of 100 depending on the grade level; the McGill Conservatorium of Music assigns 8 marks out of 100 and Mount Allison Local Center Examinations allocates 6 out of 100 marks. The Associated Board of the Royal Schools of Music piano exams, an international conservatory, also considers sight-reading to be of great importance giving it 21 marks out of 150, 14% of total marks. Music reading holds such weight in the Western musical community that it is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

also included in many music competitions through sight-reading and quick study classes (www.kiwanismusicfestivallondon.com; www.mta.ca; www.ottawakiwanismusicfestival.com). Furthermore, for such music competitions, a participant must have competed in one of the reading classes in order to be eligible to win a trophy.

Music reading continues to be seen as an important skill in the advanced stages of music training. In many universities, auditions to enter the music program require students to sight-read musical material. Auditions for the University of Alberta (www.uofaweb.ualberta.ca), Carleton University (www.carleton.ca), and the University of Western Ontario (www.music.iwo.ca) include a sight-reading test. At the University of British Columbia's School of Music, both auditions to enter the music program and term examinations include a sight-reading component (www.students. ubc.ca.). Auditions to the Boston College of Fine Arts (www.bu.edu), the University of South Carolina School of Music (www.music.sc.edu/) and to the Yale School of Music (www.yale.edu/music/) also include sight-reading exercises. At the Julliard School of Music, the ability to sight-read is among the nine criteria for acceptance to the school (www.julliard.edu). This is subsequently reinforced throughout the degree program with sight-reading and accompaniment courses (Craige, 1993).

The importance placed on music reading and sight-reading in piano method books and textbooks as well as their inclusion in conservatory exams, music competitions, and university music programs clearly displays the significance of these skills in a music education context.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.1.3. Difficulty of Music Reading

The importance of music reading is well recognized, but the development of music literacy is known to be a very long and arduous undertaking (Anderson, 1981; Hahn, 1985). The process of decoding musical notation requires both horizontal and vertical scanning of the musical score[1] and is therefore said to be more challenging than text reading, which necessitates only horizontal visual movement (Goolsby, 1994b). This process is complicated by the need for simultaneous processing of pitch, rhythmic and expressive elements, all in real time and at a specific speed. Furthermore, in addition to requiring cognitive comprehension, music reading also demands the production of a kinesthetic response on the musical instrument highlighting the fundamental requirement that the musician translate musical notation into well-controlled gestures. For all instruments, the reading of notation is a challenging endeavor. However, for reading at the piano, the performer has the additional constraint of orientation on the instrument, the pianist needing at times to remove his gaze from the score in order to look at the keyboard (Lehmann & McArthur, 2002). Pianists also generally read on a grand staff (two staves), one for the right hand and the other for the left, requiring the performer to not only read all previously mentioned musical notation elements but also process two independent lines simultaneously. A final complication with reading at the piano is the use of pedals, the damper and the *una corda*, which is often required for performance on the instrument. The complexity of this skill is further put in evidence by the poor results of students accomplishing a sight-reading task. Errors in pitch, rhythm, dynamics, articulation, and timing are commonplace (Jaarsma, Ruijssenaars, & Van Den Broek, 1998; Whitlock, 2002). These are often attributed to poorly developed reading skills

[1] When reading music on a single staff notes are not only spaced horizontally according to timing but also vertically on the staff according to pitches. Furthermore, musicians need to interpret dynamic and articulation markings which necessitate "vertical peripheral vision" (Goolsby, 1994b).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Bamberger, 1999; Gudmundsdottir, 2003; Gunter et al., 2003; Lehmann & Ericsson, 1993; Lehmann & McArthur, 2002; McKenzie et al., 1986). It can therefore be said that proper training in music reading is a necessary but arduous component of music education.

1.1.4. Need for Further Research

Even though music reading is known to be a central and challenging component of musical learning, the research on its development is not only sparse, but devoid of a theoretical foundation. In *Exploring the Musical Mind*, Sloboda (2005) states that research in the field of music reading has been limited and restricted by the common misconception that music reading is a skill which develops naturally within the learner, a view that can easily be corroborated by looking at key handbooks in music education and music psychology: *The New Handbook of Research on Music Teaching and Learning* (Colwell & Richardson, 2002), a 1222-page book, and the *Handbook of Music Psychology* (Hodges, 1996), a 589-page book, do not provide any section on music reading, and *The Psychology of Music: Cognition and Perception* (Deutsch, 1998) allocates only five pages of her more than 800 page book to sight-reading, and no section is specifically dedicated to music reading. In addition, within the existing research on music literacy, very few models or theories have been proposed (Hodges, 1992), and the models that have been developed, such as those by Kopiez and Lee (2006), Udtaisuk (2005), and Wolf (1976), are more explanatory and speculative than grounded in experimental research. Hodges (1992) writes that "in music there is no theory devoted specifically to an explanation of music reading: thus, the bulk of the research appears to be devoid of a theoretical underpinning" (p. 469). According to Dodson (1983), music reading is a complex process, the nature of which is currently not fully understood by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

music researchers. Scripp (1995) adds that little is known about the development of music reading skills. However, in order to develop further models and to pursue studies on music reading it is first necessary to better understand how music reading and sight-reading performance can be evaluated and measured.

1.2. Review of Literature

Researchers, pedagogues, and teaching institutions have used and implemented diverse methods to assess a musician's sight-reading performance. These methods could be grouped under three categories: tests and scales to measure sight-reading or sight-singing performance, quantification of errors in sight-reading performances, and evaluation by expert examiners (i.e. qualitative assessment by experienced examiners or music teachers) (see Table 2 for the review of literature organization).

Table 2. Review of Literature Organization

Sight-reading Assessment Methods	Measurement by Tests and Scales	Belwin-Mills Singing Achievement Test
		Watkins-Farnum Performance Scale
	Sight-Reading Error Quantifications	Eaton's Grading Instructions
		Gilman's Scoring Algorithm
		Gudmunsdottir's Error Classification
		Salis' Error Categorization
	Evaluation by Expert Examiners	

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.2.1. Measurement by Tests and Scales

There are very few tests of musical performance in comparison with the ever-growing number of music aptitude and achievement tests. D. A. Hodges, a leading expert in the field of music education, has written considerably on the assessment of performance skills. In his *Handbook of Music Psychology* (1996), he dedicates an entire chapter to a review of tools for evaluation and measurement in music. He explains that the assessment of performance requires the processing of several simultaneous psychomotor abilities which necessarily makes its assessment a very complicated task requiring immediate judgment. He provides valuable information on existing performance assessment scales and tests of which a small number have been developed for the assessment of sight-reading performance. The author also confirmed that no new tests have been published since he wrote his book (Hodges, D. A., personal communication, March 3, 2007). The Belwin-Mills Singing Achievement Test and the Watkins-Farnum Performance Scale, two measurement tools used in music education research, are therefore very important as they are among the only tests or scales available to measure reading skills. Nevertheless, it is important to note that such tests, though objective in nature, are influenced by the subjectivity of the varying judges' assessment (Sloboda, 2005).

1.2.1.1. Belwin-Mills Singing Achievement Test

The Belwin-Mills Singing Achievement Test (Bowles, 1971) was developed by Bowles in 1971, as a commercial venture for the Belwin Publishing Company, and was designed to test sight-singing ability from grade 5 to college seniors. This test has been used in music reading research in studies by Goolsby (1989), to evaluate twenty-four musicians'

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sight-reading ability, and Goolsby (1994b) to evaluate two musicians' sight-reading ability. However, Colwell (1991), in his review of music achievement tests, states that this test seems to have no clear purpose and that it was not meticulously developed. Hodges (1996) in his *Handbook of Music Psychology* agrees that this test was not properly standardized, there is no data reliability nor any testing norms and, therefore, this test could not be recommended with any degree of assurance. This criticism had also been previously made by George (1980) in his chapter on the *Measurement and Evaluation of Musical Behavior*.

1.2.1.2. Watkins-Farnum Performance Scale

The Watkins-Farnum Performance Scale (Watkins & Farnum, 1954), which has commonly been used in music reading research (Boyle, 1970; Elliott, 1982a, 1982b; Gaynor, 1995; Goolsby, 1994a; Gromko, 2004; MacKnight, 1975; McPherson, 1994), was originally developed in 1942 by John Watkins as a standardized evaluation scale for the cornet and it was called *Objective Measurement of Instrumental Performance*. This cornet-based scale was then adapted by Dr. Farnum in 1954 for all band instruments. This test was named the Watkins-Farnum Performance Scale (WFPS). Since this test had much success, Farnum published a similar scale for string instrumentalists, the Farnum String Scale (1969). This latter test has, however, never been used in music reading research. Boyle (1992), states that the WFPS and the Farnum String Scale are the only published and readily available performance measures.

Watkins' *Objective Measurement of Instrumental Performance* (1942) was developed as part of a doctoral dissertation. Watkins, from a list of 110 different cornet methods and 45 instrumental class methods with cornet parts, selected twenty-three of the most important

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

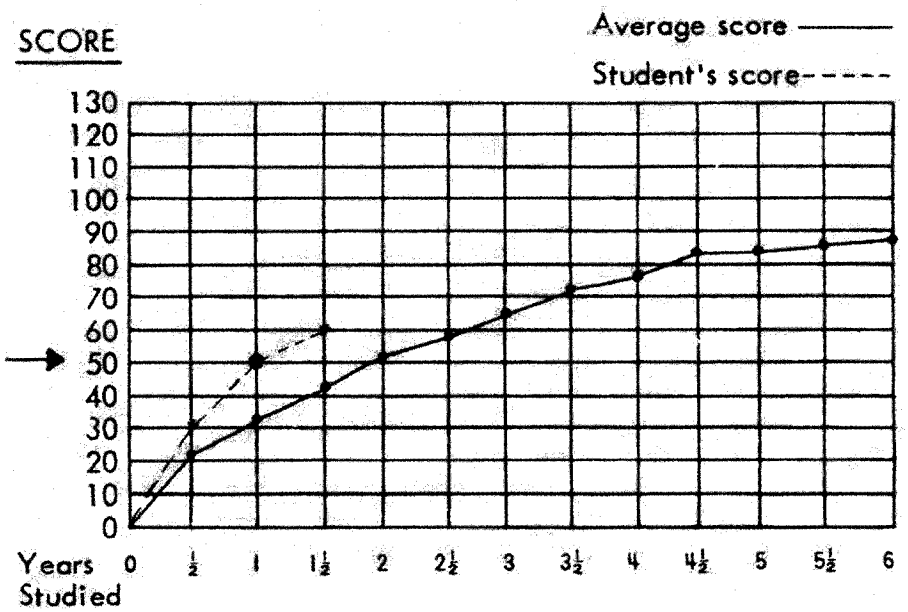
methods according to private and classroom cornet teachers throughout the United-States. The introduction of musical symbols in the twenty-three selected methods was assessed in order to compose four forms of 16 exercises each of increasing difficulty. These were administered to 105 cornet students, at which point the difficulty and performed errors in each exercise were recorded. From these four forms, two forms of fourteen exercises, each of equivalent difficulty, were constructed and named Form A and Form B. The increment of difficulty between each exercise was determined to be equal by consulting with music instructors. In preliminary testing with 105 subjects the correlation between Forms A and B was .982. Music instructors then administered the final version of Form A to 153 cornet students of various abilities; 71 of these students also took Form B. The final reliability of both tests mounted to .953. The test's validity was determined by asking the teachers whose pupils participated in the study to rank their students on their sight-reading ability and correlating these rankings to the students' performance on the test. The correlation between the teachers' ranking and the test results mostly fell in the range of .80 and .90. The student participants had also been asked questions in relation to their musical training. The highest relationship existed between the time the student first began study and the date of the test which correlated at 80.4%. Watkins' concludes by stating that this study presents a growth curve which teachers can use to compare the rates and progress for individual students.

The *Watkins-Farnum Performance Scale* (WFPS) (Watkins & Farnum, 1954) is an adaptation by Dr. Stephen E. Farnum of Watkin's *Objective Measurement of Instrumental Performance* (Watkins, 1942). Farnum states in the introduction to the WFPS that Watkins' test masters the difficulty of an objective evaluation of musical performance. He also notes that the rhythm patterns' grading of difficulty, which he deems to be the test's "most

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

noteworthy achievement" (Watkins and Farnum, 1954, p. 4), can be applied to other band instruments. Farnum therefore developed an adaptation of Watkins' cornet scale for all band instruments by (1) transposing the exercises to keys suitable to the tested instruments, (2) keeping the exercises' notes within the range of the instruments, and (3) not exceeding the difficulties and limitations of the instruments. In order to assess the two forms of the WFPS for reliability, Farnum had 71 students take both forms of the test. Farnum then validated his test by comparing the examinees' scores on the test with their sight-reading ability ranking, which was assigned by their instructor based on their "technical ability and their musicianly performance" (p. 5). Farnum also provides a chart of average scores for each half year of the subjects' six years of examinations (see Figure 1).

Figure 1. Watkins-Farnum Performance Scale: Average Scores



(Watkins & Farnum, 1954, p. 9)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In Watkins' study (1942), it was found that the relationship between the duration of the subjects' musical training and the date of their testing was extremely high. For this reason, it was deemed to be the best grading system. Thus, Farnum provides a grading chart which was designed to simplify the determination of letter grades for the students' sight-reading performance in relation to their score on the test and the number of years of study (see Figure 2).

Figure 2. Watkins-Farnum Performance Scale: Grading Chart

GRADES FOR GROUP 1 INSTRUMENTS

Years	½	1	1½	2	2½	3	3½	4	4½	5	5½	6
A	35	50	62	70	77	83	88	90	92	94	96	98
B	25	40	48	55	61	66	70	74	78	82	84	86
C	15	30	35	40	45	50	54	58	62	65	67	69
D	5	15	25	30	35	40	44	47	50	52	54	56

GRADES FOR GROUP 2 INSTRUMENTS

Years	½	1	1½	2	2½	3	3½	4	4½	5	5½	6
A	25	40	52	60	67	73	78	80	84	88	90	92
B	15	30	38	45	51	56	60	64	68	72	74	76
C	5	15	25	30	35	40	44	48	52	55	57	59
D	0	5	15	20	25	30	34	37	40	42	44	46

Sample - Clarinet score is 68 at 4 years, grade - C
Trombone score is 50 at 2 years, grade - B

(Each score is the minimum in each grade classification)

(Watkins & Farnum, 1954, p. 9)

However, Farnum found that some instruments' scores were averaging lower after one year of study than others. Instruments were therefore placed in two groups according to grading systems: Group 1: Alto Clarinet, Bass Clarinet, Clarinet (Sop.), Baritone, Cornet,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Flute, and Saxophone; and Group 2: Bassoon, Drum, French Horn, Oboe, Trombone, and Tuba. In a recent study of the WFPS by Haley (1999), in which she compares this scale to the Clarinet Performance Rating Scale (Abeles, 1973) and applies both scales to the Rasch Item Response Theory Model (Snyder & Sheehan, 1992), this test was deemed to have a high degree of validity.

This test has been used by various researchers to assess musicians' sight-reading performance. Elliott (1982a) used the WFPS to analyze the sight-reading errors made by a group of 30 randomly selected undergraduate wind instrumentalists. For each subject the tempo was introduced with a metronome which was turned off once they started sight-reading. Subjects were allowed to pause 15-20 seconds between exercises. The audio recording of their performance was then scored independently by two graduate music education students; an average of both scores was then used for each subject's final score. The purpose of this study was to identify and categorize the errors which were placed in the following categories: pitch, rhythm, expression, articulation, and other. In conclusion, it was determined that the most prevalent type of error made was within the rhythm category. However, the difficulty of rhythm reading while sight-reading was only present for the poor readers. The findings within the pitch category indicate that the leading cause of these errors was carelessness. In the same year, Elliott (1982b) conducted another experiment in which he attempted to find correlations between sight-reading ability and the following seven predictor variables: technical proficiency, rhythm-reading ability, sight-singing ability, cumulative grade point average, cumulative theory grade point average, cumulative performance jury grade point average, and major instrument grade point average. Thirty undergraduate wind instrumentalists were asked to take the WFPS, a technical proficiency

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluation, a criterion sight-singing test, as well as a rhythm-reading test developed by Boyle (1968) that was based on the WFPS. The scores from the four tests were then compared to the students' cumulative averages. It was determined that the strongest relationship exists between students' ability to sight-read and their ability to read rhythms. Performance jury scores and rhythm reading proficiency combine to make the best predictor of sight-reading ability in wind instrumentalists.

The WFPS was then used by Thompson (1985) to assess the source of individual differences in sight-reading. The sight-reading performance of 30 flutists between the ages of 17 and 31 and having been playing five to 18 years was indexed by the number of measures correctly played while they undertook the WFPS. Subjects were asked to "play each piece as quickly and accurately as possible" (p. 347). The audio recording of the WFPS performance was used for scoring; measures were marked as incorrect if subjects either played a wrong pitch or rhythm. The results of this test were then compared to the subjects' eye-performance span and their capacity to immediately recall briefly presented musical excerpts. These sight-reading factors were found to be highly correlated. The WFPS was also used in 2002 by Lafferty to assess the correlation of elements such as gender and appearance, on the performance evaluations of examiners. Eight college level clarinetists were recorded while performing the sixth exercise of both Form A and Form B. These recordings were then evaluated by 105 woodwind adjudicators from the Ohio Music Education Association (OMEA). Lafferty found that adjudicators were affected by visual elements in their assessment, providing significantly different ratings for identical performances. These briefly reviewed studies suggest that the WFPS is a well known and important tool in music education research.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.2.2. Sight-Reading Error Quantifications

Certain studies have identified and quantified the errors made by musicians while reading musical notation in order to assess sight-reading performance (Eaton, 1978; Gilman, 2000; Gudmundsdottir, 2003; Salis, 1977). Some researchers record the performances and then use the audio recordings for grading (Eaton, 1978; Elliott, 1982b; Salis, 1977). Others employ MIDI data of sight-reading samples from which they obtain a visual representation of the performance to facilitate the quantification process (Gilman, 2000; Gudmundsdottir, 2003). In such research, parameters such as timing accuracy, dynamics, articulation, and pitch accuracy are presented in computer-generated charts drawn from retrieved MIDI data of musical performances. The performed errors are then located and measured in terms of frequency, placement, and type. In this section, four classification and quantification procedures that have been developed for the assessment of music reading performance are reviewed: Eaton's (1978) grading instructions, Gilman's (2000) scoring algorithm, Gudmundsdottir's (2003) error classification, and Salis (1977) error categorization.

1.2.2.1. Eaton's Grading Instructions

Eaton (1978) assessed the correlation of sight-reading skills with previous training, note-reading skills, psychomotor skills, and memory in 73 pianists. Subjects were recruited among the pianists and organists at a music college. For the sight-reading component of this study four pieces were selected from published piano literature. During the testing, subjects were allocated 30 seconds to review each piece before commencing to play. Performances were tape recorded. Three faculty members, at the same College where subjects were recruited, graded the audio recordings of the sight-reading performances. Judges were given

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

grading sheets to assess the performances. Eaton developed a specific grading procedure for the measurement of the sight-reading performances (see Table 3). Errors were divided into two categories: note errors; and rhythm and tempo errors. In the note error category, judges were instructed to circle wrong notes, notes within chords which were incorrect, missing notes, as well as the re-striking of notes or chords. All wrong notes were given one mark. In the rhythm and tempo error category, judges were instructed to give one mark for short hesitations, two marks for long hesitations, one mark for change of pulse, one mark for incorrect rhythm, one mark for failing to observe a fermata, and three marks for failing to observe the indicated tempo.

The sight-reading grades were then compared to the grades achieved for the note naming tasks, psychomotor tasks, and memorization tasks. It was determined that the strongest relationship is between psychomotor skills and sight-reading skills and the weakest relationship is between memorization skills and sight-reading skills. To our knowledge, this method for assessing sight-reading has not since Eaton's (1978) study been used either in music reading research or in music education contexts.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3. Eaton's Grading Instructions (Eaton, 1978)

Note Errors	1. Circle the note involved in any of the following types of errors.	a. wrong note
		b. wrong chord (circle each note that is incorrect)
		c. missing note
		d. re-striking (circle the note or chord)
	2. Enter the total number of "circle" in the space indicated for note errors.	
Rhythm and Tempo Errors	1. Use the following code for indication of rhythm and tempo errors. Place the code letter directly above the error for each staff.	h = short hesitation (one pulse or less)
		H = long hesitation (more than one pulse)
		Ch = change of pulse (i.e. faster, slower,  to  ,  to  , etc.)
		R = incorrect rhythm in relation to basic tempo (i.e.,  too slow,  too fast,  not held, etc.)
		 = hold omitted (circle fermata)
		Allegro = circle tempo marking if not approximate (i.e., much faster or slower than indication)
	2. Using the following point scale, total the rhythm and tempo errors and enter in the space indicated.	h = 1
		H = 2
		Ch = 1
		R = 1
		 =1
		Allegro = 3

1.2.2.2. Gilman's Scoring Algorithm

In order to evaluate a model of eye-movements during sight-reading developed by Kinsler and Carpenter (1995), Gilman (2000) undertook ten experiments using eye-tracking technology to assess reading performances. In order to classify the subjects according to their sight-reading ability Gilman developed a scoring algorithm[2] to measure subjects' performances objectively. To check the reliability of the algorithm, Gilman asked an expert

[2] Algorithm is defined by the *Dictionary of Psychology* (Colman, 2006) as a systematic procedure for finding the distinct results to a problem. Though this definition could also be applied to the other sight-reading error quantification methods, algorithm will only be used to describe Gilman's method as she labeled it as such.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

adjudicator to rate the same subjects' performances. This algorithm was found to be a good method for the objective assessment of sight-reading performance. Gilman's scoring algorithm was organized into three categories: pitch, rhythm, and extra notes (see Table 4). These were coded separately. For the pitch grades, one point was given for each note performed which were accounted for in the musical score, 1/12 of a point was then subtracted for every semi-tone between the intended pitch and the one performed. For the rhythm grade, one point was given to each note and rest performed which were accounted for in the musical score, ¼ of a point was then deducted for each semi-quaver (sixteenth-note) beat between the performed note or rest and the intended note or rest in the musical score. For each semi-quaver beat of added notes or rests, ¼ of a point was subtracted from the rhythm grade. Rhythm and pitch grades were then added together and divided by the highest grade possible.

Gilman found that sight-reading quantification with this scoring algorithm determined grades that were consistent between both pieces and highly correlated with the expert's evaluation. This algorithm was, therefore, found to be a good means of objectively assessing sight-reading performance. This scoring algorithm has since been used by Gilman, Underwood, and Morehen (2002).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4. Gilman's Scoring Algorithm (Gilman, 2000)

Pitch Score	One point is given to each note played (apart from extra notes).	
	1/12 of a point is then deducted for each semitone between the note played and the corresponding note in the score.	Example: If the note in the score was a C but the note played was a C sharp, the subject scored 11/12 of a point; if the note played was a B flat, the subject scored 10/12 of a point.
Rhythm Score	One point is given for each note played and for each rest accounted for (apart from extras).	
	1/4 of a point is deducted for each semiquaver (sixteenth note) beat between the duration of the played note/rest and the duration of the corresponding note/rest in the score.	Example: If the note in the score was a quaver (eighth note) and the note played was a semiquaver, the subject scored ¾ of a point. If the note played was a crotchet, the subject scored ½ point.
Extra Notes	Duration of extra notes and rests were added together and ¼ of a point deducted from the rhythm score for each semiquaver beat.	Example: A pianist who makes 3 attempts at performing a note should score less than a pianist who performs a note correctly on one attempt. Similarly, a pianist who pauses (rests) before performing a note correctly should score less than a pianist who performs a note correctly without pausing.

1.2.2.3. Gudmundsdottir's Error Classification

Gudmundsdottir (2003) analyzed the music reading errors of young piano students in terms of the types of errors, the frequency of errors, and the location of errors. This study also looked at the effect of the subjects' age, and the differing complexities of exercises on the errors made. For this experiment, the performances of three unfamiliar musical pieces by 35 young piano students' were analyzed. Graphic reproductions of the sight-reading performances were used to identify the errors in relation to pitch and chord errors, this was then followed with a timing analysis. Gudmunsdottir divided the pitch category into three sections: redundant pitches, omitted pitches, and erroneous pitches which was further

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

categorized as contour-preserving and contour-violating errors. Chord errors were divided into five categories: redundant chords, switched chords, out-of-chord errors, within chord errors, and omitted pitch errors. For the timing analysis, performances were assessed in relation to total length and pauses were then assessed in relation to their length and frequency. Since subjects were not given a fixed performance tempo, rhythm was assessed by comparing the ratio between note length (i.e. was the difference in quarter-notes and eighth-notes closer to 2:1 or 1:1) (see Table 5 for a detailed explanation of the different types of errors). No guidelines are provided for the grading of performances, this study solely assessed errors for explanatory purposes.

Errors were then compared to assess the type of problems commonly faced by young readers. It was found that young piano students made significantly more errors in the left hand portion of the performance and that erroneous pitch errors were the most common type of error. It was also found that young children perform the pitches to the detriment of rhythm and timing accuracy. To our knowledge, this error classification has not been used since Gudmunsdottir (2003) in either research or education.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5. Gudmunsdottir's Error Classification (Gudmunsdottir, 2003)

<table>
<tr><td rowspan="5">Pitch Errors</td><td colspan="2">(1) Redundant pitches, correct pitches which are wrongly repeated, as if the student is hesitating.</td></tr>
<tr><td colspan="2">(2) Omitted pitches, notes that are not played.</td></tr>
<tr><td rowspan="3">(3) Erroneous pitches; incorrect notes that did not match the note indicated in the musical score. These were marked as 'self-correction' when followed by the correct note. These are divided into two sub-categories:</td><td>(1) Contour-preserving errors, which were in the same direction as the accurate pitch.</td></tr>
<tr><td>(2) Contour-violating errors, which were in a different direction than the correct pitch.</td></tr>
<tr><td rowspan="5">Chord Errors</td><td rowspan="5">Chords are deemed wrong when any one or more of the notes are played incorrectly or missing. They are also marked as incorrect when they are repeated more than indicated in the score. After having found a chord to be incorrect, further analysis was undertaken to find the nature of the error. These are placed in five categories:</td><td>(1) Redundant chords, when the chord was played more times than was indicated in the score.</td></tr>
<tr><td>(2) Switched chords, neighboring chords played in reverse order.</td></tr>
<tr><td>(3) Out-of-chord errors, chords where one or more of the pitches are incorrect and do not belong to the intended chord.</td></tr>
<tr><td>(4) Within-chord errors, chords where one or more of the pitches are incorrect but do harmonically belong to the intended chord.</td></tr>
<tr><td>(5) Omitted pitch, chords where some of the intended notes are missing.</td></tr>
<tr><td rowspan="3">Timing Analysis</td><td colspan="2">Performances were calculated from beginning to end in milliseconds. This was assessed in terms of performance length and then related to the subject's age group.</td></tr>
<tr><td colspan="2">The next type of analysis made within the timing category was to evaluate the number of pauses made in a performance. A pause was counted when it was of equal length or longer than the average duration of the two notes before and the two notes after. The frequency of pauses was then calculated in relation to age group. The mean length of pauses was also measured in milliseconds.</td></tr>
<tr><td colspan="2">The subjects were permitted to choose their own performance tempo, the duration of notes were measured by determining whether the ratio between quarter-notes and eight-notes was closer to 2:1 or 1:1. The reasons given for this method of assessment is that children have great difficulty following the metronome beat and younger readers tend to focus on the pitch to the detriment of timing and rhythm.</td></tr>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.2.2.4. Salis' Error Categorization

Salis (1977) studied cognitive variables' relation to the reading of piano music. A music reading test was developed in order to classify participants according to their sight-reading ability. Though Salis named her assessment method a test, it is clearly a method of sight-reading error quantification. For the classification of 26 subjects, either music majors in college or professional pianists, 45 music excerpts were selected from the keyboard literature of Bach, Mozart, Beethoven, Brahms, Debussy and Ravel. From the chosen excerpts 1/3 were deemed easy in reading level, 1/3 medium, and 1/3 hard by three professional musicians. This test evaluates musicians solely on pitch and rhythm errors and does not incorporate an analysis of "artistic rendition" (p. 47). Subjects were given 30 seconds to preview the piece. If a subject encountered a piece he/she had already studied a replacement piece was given. For the scoring of performances, the experimenter used the audio recording of the sight-reading performances and only the excerpts of medium and hard difficult levels were assessed. The errors were divided into two categories: tonal errors, and rhythmic errors (see Table 6). Tonal errors included omitted notes, omitted key signature accidentals, other omitted accidentals, wrong pitches, inserted notes, and misread clefs. Rhythmic errors included wrong notes or chord values, repeated notes or chords, omitted ties, indefinite prolonging, and incorrect synchronization of the two hands. One point was given for each type of pitch error and each type of rhythm error; these were accumulated separately to provide a total pitch score and a total rhythm score.

This study found that IQ, word reading skills, and the perception of tachistoscopically controlled words are strongly correlated to music reading skills. This study also found that good music readers have a greater ability to perceive the relation within and between musical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chunks. Furthermore, timing seems to be a determining factor between good and poor music readers, good readers being more efficient in their performances. To our knowledge, Salis' error quantification method (1977) has not been used since in either research or education.

Table 6. Salis' Error Categories (Salis, 1977)

Some General Conditions	1. Each excerpt of each style was tallied separately.
	2. Tonal and Rhythmic errors were tallied separately.
	3. Errors for chords and single notes were tallied separately.
	4. Errors for right and left hands were tallied separately.
	5. Octaves, whether by themselves or as part of a chord, are treated as 1 note.
Tonal Errors	1. Omission: 1 for each notes omitted.
	2. Key signature accidentals: 1 for each note with a key signature accidental omitted.
	3. Regular accidentals: 1 for each note with an inserted accidental omitted (accidentals not part of the key signature).
	4. Letter change: 1 for each note with a letter change.
	5. Inserts: 1 for each note inserted into the music (notes that were extra and not simple letter changes).
	6. Wrong clef: 1 for each note misread because of clef change.
Rhythmic Errors	1. Wrong value: 1 for each note or chord with wrong note value.
	2. Repeats: 1 for each time a note or chord was repeated. Repeats that involved both hands were counted as 1 and .5 error attributed to each hand.
	3. Omission of tie: 1 for each omission. Notes of a chord tied together were counted as 1 omission.
	4. Indefinite prolonging: 1 for each place in the music where the beat flow was interrupted by prolongation not recognizable as a note value change. Since both hands usually stopped at such places, .5 error was attributed to each hand.
	5. Incongruities: 1 for each place where the synchronization of the two hands was incorrect, i.e., playing a right hand note simultaneously with a left hand note when it should have been played after the left hand note. Missing rests in one hand or the other also caused this to happen.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.2.3. Evaluation by Expert Examiners

The final method of sight-reading performance assessment to be discussed here is evaluation by expert examiners. This type of assessment is used in music reading research, as well as in music examination, and music competition settings.

Many studies into the cognitive processes of music reading have made use of expert evaluations to rate performances. Lehmann and Ericsson (1993) had two "experienced jurors" (p. 189) evaluate the audio recording of 16 university level piano student's performances of two accompaniment pieces. Rogers (1996) used a judge to assess the reliability of the testing procedure in the evaluation of the sight-tapping of two musical rhythms by a group of 134 first and second grade students. Furneaux and Land (1999) used grades from the Associated Board of the Royal Schools of Music piano exams (ABRSM) to choose their six subjects who were to be compared with two professional accompanists. In ABRSM piano exams performances are evaluated by a trained examiner, students are given a grade out of 21 marks for the sight-reading portion. In another experiment, Gilman (2000) asked one expert to evaluate the audio-recording of the performance of two pieces by 24 pianists. Levy (2001) asked three judges, using video-taped recordings, to evaluate and transcribe the errors made in the sight-reading performances of 11 band students. Wöllner, Halfpenny, Ho, and Kurosawa (2003) had three expert listeners (experienced music lecturers) judge the sight-singing performance of two musical excerpts by 20 undergraduate and postgraduate music students in relation to continuity, fluency, and quality.

In piano examination settings across Canada, sight-reading performance grades are based on expert examiners' evaluations. Consulting senior examiners at two major conservatories (The Royal Conservatory of Music of Toronto and Conservatory Canada),

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

made it clear that no scales of evaluation or specific marking criteria are utilized in such assessment practices. This is also the case in competitions, such as Kiwanis music festivals, where professional adjudicators are asked to judge sight-reading performances. It is interesting to note that in settings where evaluations are determined by expert examiners, assessment reliability has often been found to be very low (Thompson, Diamond, & Balkwill, 1998). Nevertheless, it is important to take evaluation by expert examiners into account when discussing sight-reading performance assessment since it is the most frequent method used in educational contexts. Furthermore, it has also been used in music reading research to rate subjects according to their sight-reading performance (Lehmann & Ericsson, 1993; Levy, 2001; Rogers, 1996; Wöllner, Halfpenny, Ho, & Kurosawa, 2003) and to validate tests, scales, and error quantification measures (Gilman, 2000; Watkins, 1942).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.3. Purpose of the Study

This study focuses on the measurement, quantification, and evaluation of reading and sight-reading skills. Knowing how to assess these skills is essential in order to conduct further research on music reading and eventually make possible the creation of cognitive models for reading and sight-reading development, in addition to providing new insight into this complex skill and guidelines for teaching it. Hodges (1980) states that "the evaluation of performance achievement is an essential and a pervasive aspect in the process of music education, yet compared with other types of published music tests, relatively little effort has been expended toward its measurement" (p. 334). Music reading and sight-reading are important aspects of performance, and thus far, guidelines for the evaluation and measurement of reading accuracy in music performance have not been standardized within the research. It is also important to note that no sight-reading measurement scale has been developed for assessing pianists.

Research that discusses the assessment of music reading used very different methods, including sight-reading tests and scales (Elliott, 1982a, 1982b; Lafferty, 2002; Thompson, 1985), quantification of sight-reading errors (Eaton, 1978; Gilman, 2000; Gudmundsdottir, 2003; Salis, 1977) and evaluation by expert examiners (Furneaux & Land, 1999; Lehmann and Ericsson, 1993; Levy, 2001, Rogers, 1996). These methods have, however, never been compared to determine if they provide analogous assessments of reading accuracy in performance. When assessment methods used in research vary considerably, it is difficult to provide reliable and analogous data. Consequently, comparing studies that have used different standards and methods of assessment is a challenging task. Therefore the following question will be addressed: Do the measurement, quantification, and evaluation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

methods used by researchers to assess music sight-reading provide comparable analyses and equivalent results?

The purpose of this project is twofold. First, we will develop a sight-reading measurement scale for pianists that will be adapted from the Watkins-Farnum Performance Scale for woodwind players. Second, we will experiment with the three sight-reading assessment methods introduced so far, sight-reading performance tests and scales, sight-reading error quantifications, and evaluation by expert examiners, to compare their assessment procedure and to determine whether these methods provide equivalent analyses of sight-reading performance in pianists. It is hoped that this study will further our understanding of the varying methods implemented for the assessment of sight-reading performance in music reading research.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2

2. Methodology

For the purpose of this study, three methods of sight-reading performance assessment have been selected for comparison: 1) sight-reading tests and scales, 2) sight-reading error quantifications, and 3) evaluation by expert examiners. In the first category, two tests and scales were considered. For obvious reasons the Belwin-Mills test is not satisfactory: 1) it was developed to measure sight-singing ability and would therefore require a large amount of revision and adaptation for use with piano sight-reading performance; and 2) this test was not meticulously developed and it was found to have no data reliability by Colwell (1991), Hodges (1996) and George (1980). However, the WFPS was chosen for its high level of reliability and its well-detailed procedure, as well as its clear, well-organized, and thoroughly developed measurement criteria, in addition, it is more commonly used in music education research settings than the Belwin-Mills. Since this test was devised for the assessment of woodwinds, an adaptation is necessary so it can be used in the context of piano playing. New musical excerpts will need to be selected using the grand staff and taking into consideration the difficulties of piano playing. Its correction criteria will also require revisions for the assessment of pianists.

For the quantification of sight-reading errors, four systems were considered. It was decided to only keep one of the four approaches. After a good review, Eaton's (1978) and Salis' (1977) grading procedures, though they were developed for use with pianists and provided relatively thorough methods of measurement for the assessment of sight-reading, were not kept since to our knowledge in the 30 years since their conception they have never been used and seldom mentioned in other research projects to date. Furthermore, Eaton's

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(1978) error quantification method does not seem to have been validated and despite the fact that Salis' (1977) grading procedure is said to have "face or content validity" (Salis, 1977, p. 47), it does not appear as though it was validated by any methodical or standardized procedure. Gudmunsdottir's error analysis, even though it was developed for use with young piano students and provides a well described and clear error characterization, was not kept since it does not provide a grading method. Her error characterization was developed to study the type and frequency of errors made by young children while reading musical notation and it did not provide an error grading system. Since the purpose of our study is to compare assessment methods, it was determined that this error classification would not be useful for our project. Gilman's (2000) scoring algorithm provides a well constructed and complex method of quantifying sight-reading data. It was also the only error quantification method validated where the results were compared with an expert examiner's evaluation. This grading method has also been used in one other research project by the same author to assess sight-reading performance (Gilman, Underwood, and Morehen, (2002). Furthermore, it is a more recent and up to date method of assessment than both Eaton (1978) and Salis (1977). This scoring method was also developed for analysis of piano sight-reading performance which makes it appropriate to our project.

For the evaluation by expert examiners component of the analysis, three expert examiners were chosen on the basis of their experience and reputation in the area of evaluating and teaching piano performance. Their evaluations will be compared with the two other sight-reading performance assessment methods.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.1. Developing a Scale

Three steps were required for the assessment of piano sight-reading performance with an adaptation of the WFPS: 1) finding appropriate musical stimuli, 2) adapting the grading system used in the WFPS, and 3) adapting the grading chart.

2.1.1. Finding Musical Stimuli

New musical material was needed as the WFPS provided exercises for wind instruments only. It was decided that the composition of new musical material would be the best way to ensure that the stimuli were unknown to subjects. The new musical excerpts were therefore commissioned to the composer Mary Gardiner. Two elements needed to be addressed: 1) the number of exercises in the scale and increment of difficulty of the exercises to be composed, and 2) the number of measures in each exercise.

The number of exercises in our scale and the increment of difficulty between each exercise was the first element needing modifications. The WFPS is a series of 14 exercises which are of increasing difficulty, from very simple to fairly complicated. For the new piano sight-reading scale, we chose to select only 11 exercises. This conformed to the norms of the Royal Conservatory of Music (RCM) as well as those of many other music examination centers. This grading system is made up of 10 grades in addition to an ARCT level. The RCM is one of the leading piano examination centers in Canada and its grading systems have been exported to many countries around the world. This level structure was also selected as many piano teachers and students participating in our study would be familiar with the RCM's norms. Furthermore, for the evaluation by expert examiners component of our project, selected adjudicators had prior knowledge and experience with this grading system.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The RCM's syllabus provides a clear description for the increasing difficulty of exercises from levels 1 to 11. Therefore, the commissioned composer was given a set of guidelines, as defined by the Royal Conservatory of Music, for the composition of each of the 11 sight-reading exercises (see Table 7). The composer was also asked to compose two exercises for each level of difficulty, which would result in two sets of equal difficulty.

The newly composed exercises were reviewed by two experts very knowledgeable in the RCM's standards for sight-reading examinations. The two experts were sent either print or electronic copies of the sight-reading pieces for grades 1 to 11 of the first set. After a careful revision of the exercises, they provided comments about the suitability of each one according to the guidelines implemented for the RCM examinations' sight-reading material. The comments received were very positive and highly praised the compositions. Comments such as: "Obviously Mary Gardiner knows students well, and all the pieces fall nicely into their designated grades in terms of difficulty... Totally suitable for the level... Lovely.... Perfect! Excellent! Delightful little piece... I love this piece! Well within Grade 7 level of difficulty..." were among the praises given to the compositions and their appropriateness for the intended level of sight-reading difficulty. The experts also provided a list of detailed criticism such as: "I find the ***sound*** of the first beat in m. 3 a little stark (F# octave) in the midst of the harmony that precedes and follows it... This tests a student's ability to comprehend and coordinate contrapuntal style. Because of the complexity of articulation and dynamics, this would seem to me to be more difficult for Grade 6 than the pieces designated for the previous levels..." This previous criticism was followed by detailed commentary and suggestions to the fingering proposed by the composer. Their observations resulted in modifications to two of the compositions, exercises six and eleven of the first form. Their

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comments were also taken into consideration by the composer for the composition of exercises 1 to 11 of the second set. The exercises for the second set were reviewed by one of the experts. These were found to be appropriate in both musicality and difficulty to RCM standards.

Table 7. Requirements for Sight-reading as Listed in the 2001[3] Edition of the Royal Conservatory of Music Piano Syllabus

Grade 1	Candidates will be asked to play a simple four-measure melody in the key of C major or F major with the given fingering. The melody will be divided between the hands and will be within the compass of the staff. The rhythm will consist of quarter and half notes.
Grade 2	Candidates will be asked to play a simple four-measure melody in the key of C, F, or G major, with the given fingering. The melody will be divided between the hands, and the rhythm will consist of half, quarter, and eighth notes.
Grade 3	Candidates will be asked to play a simple short melody in 4/4 time, hands together, in the keys of G major, or D major. The right-hand part will be in quarter or half notes; the left-hand part will be in half notes.
Grade 4	Candidates will be asked to play a simple passage in 4/4 time, hands together, in the keys of G, D, F, or A major. The right-hand part will be in eighth, quarter, or half notes; the left-hand part will be in half notes. The melody will begin and end with a triad in the left hand.
Grade 5	Candidates will be asked to play a short passage in either a major or minor key, approximately equal in difficulty to repertoire of a Grade 2 level.
Grade 6	Candidates will be asked to play a passage in either a major or minor key, approximately equal in difficulty to repertoire of a Grade 3 level.
Grade 7	Candidates will be asked to play a short composition equal in difficulty to repertoire of a Grade 5 level.
Grade 8	Candidates will be asked to play a short composition equal in difficulty to repertoire of a Grade 6 level.
Grade 9	Candidates will be asked to play a short composition equal in difficulty to repertoire of a Grade 7 level.
Grade 10	Candidates will be asked to play at sight given passages of music of contrasting styles.
ARCT	Candidates will be asked to play at sight a given passage approximately equal in difficulty to repertoire of a grade 9 level.

[3] In the Spring of 2008 a new edition of the RCM's piano syllabus was published, however, when the musical exercises for our Piano Sight-reading Scale were composed by Mary Gardiner the syllabus had not yet been released. For this reason the RCM's 2001 guidelines were used for this research project. Nevertheless, the changes to the sight-reading portion of the syllabus were minimal and the exercises developed for the purpose of this project are still relevant in the context of the new RCM piano syllabus.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The second element which was modified for the piano sight-reading scale was the number of measures in each exercise. Instead of the 16 to 36 measures of the WFPS, the Piano Sight-reading Scale exercises were significantly shorter, ranging between 4 and 18 measures. This was necessary because the guidelines set by the RCM state that for grades one and two, piano students will be asked to sight-read a four measure musical exercise. The nine subsequent exercises could then not jump significantly in the number of measures. Furthermore, most material used to practice sight-reading in the first few years of piano lessons is typically very short. In Berlin's (1986) *Four Stars Sight-Reading and Ear Tests: Daily Exercises for Piano Students*, the average number of measures per exercise is 4.49 in the grade one book, 4.51 in the grade two book, 4.63 in the grade three book, 4.76 in the grade four book, 7.89 in the grade five book, 8.57 in the grade six book, 7.81 in the grade seven book, 9.09 in the grade eight book, and 10.16 in the grade nine book. In Bennett and Capp's (1969) *Complete series of sight-reading and ear tests,* the sight-reading books developed for use with RCM examinations, the average number of measures per exercise is 3.97 in the grade one book, 4.31 measures in the grade two book, 4.25 in the grade three book, 4.35 in the grade four book, 8.17 in the grade five book, 10.24 in the grade six book, 9.18 in the grade seven book, 10.04 in the grade eight book, 10.33 in the grade nine book, 10.28 in the grade ten and eleven book. Therefore, for our piano sight-reading scale, exercises one and two contain four measures each, exercise three has eight measures, exercise four has six measures, exercise five has eight measures, exercise six and seven have nine measures each, exercise eight has eleven measures, exercise nine has nine measures, exercise ten has eighteen measures, and exercise eleven has fourteen measures, in order to conform with the RCM's norms as well as those of important sight-reading material.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The WFPS consists of two forms of equal difficulty, Form A and Form B, each of 14 exercises of increasing complexity. However, many researchers having used the WFPS for the purpose of music reading research only employed one form in the course of their studies. Elliott (1982a, 1982b) only had participants sight-read exercises from Form A of the WFPS during the course of his experiments. Studies by Thompson (1985), McPherson (1994) and Gromko (2004) also only made use of Form A of the WFPS to assess musicians' sight-reading performance. Furthermore, the purpose of this study is to adapt the WFPS for use with pianists and to use this adapted scale to compare the three predominant methods of assessment, the WFPS, sight-reading error quantification, and evaluation by expert examiners. A second form would therefore not be necessary for this research project. Although two forms were commissioned and composed, it was decided to employ only one form in the course of our study.

To conclude, the musical stimuli employed for our study consisted of eleven exercises of increasing difficulty composed according to the RCM's guidelines. Comforming with RCM norms, as well as those of important sight-reading material for piano, exercises ranged from four to 18 measures in length. After reviewing a number of studies employing the WFPS to assess participants sight-reading skills, it was decided that only one form would be used for our research project.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.1.2. Revising the Grading System

The second element of the WFPS needing to be adapted for use with pianists was the method used for grading the participants' performances. With the WFPS, exercises have a standard maximum grade (10 being the most prevalent); this grade represents approximately two-thirds of the total number of measures in the piece. Watkins (1942) chose the standard maximum grade by first asking five cornet teachers the following question: "If you were grading a student on sight-reading ability and students were given exercises of sixteen measures each which gradually increased in difficulty, about how many wrong measures would you permit and still consider that the student should receive some credit?" (p. 51). Three teachers answered about half of the bars, one said a little more than half, and the fifth teacher answered one error in each bar. To further confirm that two-thirds was an appropriate standard maximum, after having 105 cornet students take his test, Watkins assessed the number of exercises played with no errors, one error, two errors, etc, up to two-thirds of the number of measures. Since most human skills seem to be distributed evenly among the population, after having analyzed the errors made by the students, the number of measures performed correctly was "translated into standard deviations on a normal distribution curve" (Watkins, p. 57). It was then determined that two-thirds of the total measures was an appropriate standard maximum grade for each exercise.

Starting with the standard maximum grade, points are then deduced for each measure played incorrectly. A single mistake in a measure reduces the final score by one point. A maximum of one point can be deduced per measure played incorrectly. Once approximately two-thirds, the exact number varies depending on the exercise, of the measures in one exercise are performed with errors, that exercise's final grade is zero. The examinee is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

instructed to play the exercises in order, until the examiner deems the difficulty too great, on the basis of the number of errors being made, and asks the subject to stop. This point is reached when the student has scored 'zero' for two consecutive exercises. Each subsequent exercise is accorded a 'zero' by assumption.

The grading system needed to be adapted for the piano sight-reading scale. In piano sight-reading, subjects are generally required to play more than one note at a time and teachers and examiners alike would agree that a standard maximum grade of two-thirds of the total measures in the exercise is unreasonably harsh. Therefore, conforming to the method used by Watkins to establish his grading system, expert examiners with the Royal Conservatory of Music were consulted with the following question:

> If you are grading piano students on sight-reading ability and students are given exercises of four to eighteen measures, each which gradually increase in difficulty, **how many wrong measures** would you permit and still consider that the student should receive some credit?

Since pieces varied in length and teachers might have found it difficult to answer the question with a number of measures, the following five options were given as choices:

a. all measures
b. almost all measures
c. three-quarters of the total measures
d. half of the total measures
e. less than half of the total measures

Examiners were then instructed to highlight the answer which best represented their judgment of piano sight-reading evaluation and to provide comments on their choice (see

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix B for the Adjudicator Question Document). Four Royal Conservatory of Music's examiners were recruited via email. Two of the examiners answered that they would permit mistakes in all measures and still allow some credit and the other two examiners answered that they would permit mistakes in almost all measures and still allow for some credit. We therefore chose to grade each exercise according to the number of measures in the exercise (see Table 8 for exact grading measure of each exercise).

Table 8. Grading Measure for the Piano Sight-reading Scale

	Grading measure
Exercise 1	4 marks
Exercise 2	4 marks
Exercise 3	8 marks
Exercise 4	6 marks
Exercise 5	8 marks
Exercise 6	9 marks
Exercise 7	9 marks
Exercise 8	11 marks
Exercise 9	9 marks
Exercise 10	18 marks
Exercise 11	14 marks

The grading chart proposed by the WFPS (see Figure 2) also needed to be adapted. The WFPS' grading chart provides a letter grade for the students' sight-reading performance in relation to their final grade on the test and the number of years of study on their musical instrument. None of the research projects using the WFPS mention the grading chart and it can therefore be assumed that it wasn't used in such studies (Elliott, 1982a, 1982b; Gaynor,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1995; Goolsby, 1994a; Gromko, 2004; MacKnight, 1975; McPherson, 1994). In addition, this grading system provides marks for subjects from six months of instrumental instruction to six years of instrumental instruction, when in many of the research projects subjects had been studying music for more than six years. Thompson's (1985) subjects had been playing the flute from five to 18 years. Elliott (1982a & 1982b), Goolsby (1994a), and Lafferty (2002) used University and College level music students; it can probably be assumed that the majority of these subjects would have been studying their instrument longer than six years. This could explain why such research does not make use of this grading chart. Furthermore, in a later study of flutists by Thompson (1985), it was found that sight-reading ability is not related to the number of years a musician has been playing his instrument. This finding brings into question the WFPS's grading chart. Since research projects using the WFPS for evaluation purposes do not seem to make use of this grading chart, the possible fact that it is not a reliable means of assessment, and that most of our participants have studied the piano for a longer period than six years, we decided not to utilize this grading chart in our research project.

2.1.3. Adapting the Error Characterization

The error characterization used for the WFPS is clear, comprehensive, and exhaustive (see Table 7). An error coding system is used to measure the performance: (P) for pitch errors, (R) for time errors (rhythm) as well as rests, holds, and pauses, (T) for change of time errors, (E) for expression errors, and (S) for slur errors. Pitch errors include added and omitted tones, as well as notes played on the wrong pitch. Rhythm errors include all notes and rests which are not held for their correct value as well as pauses within a measure, however, pauses between measures are not considered errors. Change of time errors are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

counted in all measures played in a different tempo than the initial one chosen. Expression errors are marked for failing to perform expression or articulation markings. Failure to observe expression markings constitute as expression errors. Slur errors include slurred tongued notes, carrying a slur onto notes which should be tongued, or breaking a slur. Failing to perform a repeat also constitutes an error. The guidelines for evaluation are clearly and meticulously described for the examiner in the introduction of the WFPS (see Table 9). As these guidelines were developed for the testing of band instruments, some elements, like all discussion of the tonguing of notes, cannot be transferred to piano performance. Also, certain elements necessary in piano performance are missing from the chart, such as pedaling, coordination of hands, etc. Therefore, we had to thoroughly revise many sections to make the chart appropriate for piano assessment (see Table 10 for the Piano Sight-reading Scale's error classification).

In the Pitch Errors section of the WFPS, it is written "Fuzzy attacks or minor irregularities in pitch during the course of an extended tone are not to be counted as errors, provided most of the note has been played on the right pitch." Fuzzy attacks are a problem which can be encountered in piano sight-reading, for example when a note adjacent to the correct note is accidentally touched when playing the intended note. However, this section needed to be revised since irregularities in pitch can not occur at the piano once the note has been played. For this reason, in the piano sight-reading error classification we stated the following:

> Fuzzy attacks, more specifically when a key adjacent to the correct one is accidentally touched when playing the right key, are not to be counted as errors.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9. Watkins-Farnum Performance Scale (Watkins & Farnum, 1954)

1. Pitch Errors (P)	a. A tone added or a tone omitted constitutes an error.	
	b. A tone played on the wrong pitch.	(1) Fuzzy attacks or minor irregularities in pitch during the course of an extended tone are not to be counted as errors, provided most of the note has been played on the right pitch.
		(2) If the student strikes the wrong pitch when attacking a note but correctly fingers it and immediately adjusts the lip to the correct pitch without retonguing the note, no error is to be counted. If he fingers it wrong, for example playing F sharp and then changing immediately to F natural, an error is counted.
2. Time Errors (R) for rhythm	a. Any note not given its correct value is marked wrong.	(1) A sustained note must be held within one count of the correct beat. Thus, a whole note held for three full counts is marked wrong. If held for three counts and a little more it is considered right. It must extend over into the beginning of the fourth count. If it extends past the end of the fourth count, into the beginning of the fifth it again becomes wrong. Count to yourself and mark an error if the tone stops before you start to say the word "four" or after you have started to say the word "five."
		(2) Rule (1) above is to be applied to sustained tones, half notes, dotted halves, dotted quarters in six eight, etc. Be sure that the difference between the time any sustained tone is held and the correct time for it is less than a full beat.
3. Change of Time Errors (T)	a. If there is a marked increase in tempo or decrease in tempo, all measures played in the incorrect tempo are wrong.	(1) If the increase or decrease in tempo within an exercise is less than twelve (12) beats per minute, no errors are to be marked. Before giving the tests, practice with the metronome to determine the limits and then use your judgment when administering the test.
		(2) Increase in tempo: (a) Where the increase is in excess of twelve beats per minute, mark wrong only the measure in which the increase took place. If the increase has been gradual, mark the measure wrong in which you think it passed the twelve beats per minute mark of increase. (A return to normal tempo at any time is not counted as an error.)
		(3) Decrease in tempo: (a) Where the decrease is in excess of twelve beats per minute, mark wrong all measures played at that retarded speed. (b) If the drop in speed is continued, mark four (4) measures wrong, and then, if the student seems wedded to the slower tempo, stop him and inform him, that he is going too slow. Indicate the correct tempo with the metronome and let him start again at that point. If he again drops below the limit in tempo, say nothing but score all measures wrong which are played too slowly. (Follow the above procedure if the student fails to play exercise No. 9 in *alla breve* time.
4.Expression Errors (E)	a. Failure to observe any expression mark constitutes an error.	(1) The fact of the response, not the degree of the response, determines whether or not an error has been made. (a) Thus an increase in volume made when *f* follows *p* or *mf* indicates that the student knows the meaning of Forte and has read the symbol correctly. The fact that he has increased volume too little or too much in the judgment of the scorer shall not be counted as an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

		error. Be convinced only that the student has seen the mark, knows what it means, and indicates so by responding to it.	
		(2) Failure to observe a crescendo, decrescendo, accelerando, or *ritardando* constitutes a single error which is ascribed to the measure in which the sign originates.	
5. Slur Errors (S)	a. A slur omitted, a tongued note slurred, a slur carried onto notes which should be tongued, or a broken slur are all counted as errors.	(1) Occasionally one finds a student who has been taught to play with an especially legato tongue. Do not mark slur errors when you first discover this, but inform him it is his responsibility to distinguish between slurred and tongued notes. After that, grade him as strictly as others. If in doubt, mark as error. The burden of proof lies with the student. See that he understands that and then score rigidly.	
6. Rests (R)	a. Ignoring a rest or failure to give a rest its correct value is an error. Apply the same standards as to a sustained note (see rule 2).		
7. Holds and Pauses (R)	a. Holds written thus (fermata) should be treated in the same way as other expression marks (See section 4a, rule (1))		
	b. Pauses between measures, no matter how long, are not to be counted as errors; however, do not inform the student of this fact.		
	c. Pauses between notes within the measure are to be counted as errors.		
8. Repeats	a. Record an error if the pupil fails to make the repeat in No. 4 but score only the first rendition. In No. 7, score only the first rendition plus the second ending. Allow the student to complete the repeat but do not change the scoring or mark any new errors until the second ending	(1) If he should stop and ask you immediately whether he should repeat, answer, "Of course, play it exactly as written" and do not record an error. Failure to play the second ending on the repeat in No. 7 constitutes an error on the first measure of the second ending. In this case, tell the student to start on the second ending and finish.	

The next section which was problematic for piano sight-reading is the Slur Errors section. For this section the WFPS gives the following instructions: "A slur omitted, a tongued note slurred, a slur carried onto notes which should be tongued, or a broken slur are all counted as errors." A tongued note indicates for wind players when the tongue is used to rearticulating notes, making them sound slightly detached. This could be discussed for pianists as the articulation of notes, legato playing or the slurring of notes versus detached or staccato playing. This section was therefore rewritten as follows:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> A slur omitted, a staccato omitted, a slur carried onto notes which should be detached or staccato, or a broken slur are all counted as errors.

In the Slur Errors section, the WFPS also provides the following instructions: "Occasionally one finds a student who has been taught to play with an especially legato tongue. Do not mark slur errors when you first discover this, but inform him it is his responsibility to distinguish between slurred and tongued notes..." This applies to piano sight-reading but needed to be rephrased. Many piano teachers do not stress the importance of respecting articulation markings when sight-reading, and some students might ignore them altogether. It would therefore be important to advise students of the need to perform with the articulation indicated in the score. This section was then rephrased into the following instructions:

> Occasionally one finds a student who does not interpret articulation when sight-reading. Do not mark articulation errors when you first discover this, at the end of the exercise inform the pianist it is his responsibility to distinguish between slurred and detached phrases. For every subsequent exercise count articulation errors.

Section eight, Repeats, is not pertinent to our sight-reading scale since no repeats are found in the pieces composed for our test. This section was therefore removed altogether. Two components of piano playing which are not covered in the WFPS are pedaling and the synchronization of the two hands; therefore, two additional sections were added for the piano sight-reading classification. Section eight, Pedaling:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A student is to be marked wrong if pedal markings are not respected.

And the following specifications are given for marking pedaling errors:

(1) If a general instruction is given at the beginning of the piece to use pedaling which the student fails to observe, he is to be marked wrong only in the first measure of the piece.

(2) If pedal markings appear in individual measures, each measure with an indication to pedal which is not pedaled is marked incorrect.

A ninth section was then added for Hand Synchronization. Sight-reading at the piano requires the simultaneous performance of two separate parts, one for the right hand and one for the left hand. For this reason, in our piano sight-reading error classification it was necessary to have a section for the assessment of the synchronization between these parts. The following instructions were written for Hand Synchronization:

A student is to be marked wrong if he does not play the notes in the RH with the appropriate notes in the LH.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10. Piano Sight-reading Scale Error Classification

1. Pitch Errors (P)	a. A tone added or a tone omitted constitutes an error.	
	b. A tone played on the wrong pitch.	(1) Fuzzy attacks, more specifically when a key adjacent to the correct one is accidentally touched when playing the right key, are not to be counted as errors.
2. Time Errors (R) for rhythm	a. Any note not given its correct value is marked wrong.	(1) A sustained note must be held within one count of the correct beat. Thus, a whole note held for three full counts is marked wrong. If held for three counts and a little more it is considered right. It must extend over into the beginning of the fourth count. If it extends past the end of the fourth count, into the beginning of the fifth it again becomes wrong. Count to yourself and mark an error if the tone stops before you start to say the word "four" or after you have started to say the word "five."
		(2) Rule (1) above is to be applied to sustained tones, half notes, dotted halves, dotted quarters in six eight, etc. Be sure that the difference between the time any sustained tone is held and the correct time for it is less than a full beat.
3. Change of Time Errors (T)	a. If there is a marked increase in tempo or decrease in tempo, all measures played in the incorrect tempo are wrong.	(1) If the increase or decrease in tempo within an exercise is less than twelve (12) beats per minute, no errors are to be marked. Before giving the tests, practice with the metronome to determine the limits and then use your judgment when administering the test.
		(2) Increase in tempo: (a) Where the increase is in excess of twelve beats per minute, mark wrong only the measure in which the increase took place. If the increase has been gradual, mark the measure wrong in which you think it passed the twelve beats per minute mark of increase. (A return to normal tempo at any time is not counted as an error.)
		(3) Decrease in tempo: (a) Where the decrease is in excess of twelve beats per minute, mark wrong all measures played at that retarded speed. (b) If the drop in speed is continued, mark four (4) measures wrong, and then, if the student seems wedded to the slower tempo, stop him/her and inform him/her, that he/she is going too slow. Indicate the correct tempo with the metronome and let him/her start again at that point. If he/her again drops below the limit in tempo, say nothing but score all measures wrong which are played too slowly.
4.Expression Errors	a. Failure to observe any	(1) The fact of the response, not the degree of the response, determines whether or not an error has been

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(E)	expression mark constitutes an error.	made. (a) Thus an increase in volume made when *f* follows *p* or *mf* indicates that the student knows the meaning of Forte and has read the symbol correctly. The fact that he has increased volume too little or too much in the judgment of the scorer shall not be counted as an error. Be convinced only that the student has seen the mark, knows what it means, and indicates so by responding to it.
		(2) Failure to observe a crescendo, decrescendo, accelerando, or ritardo constitutes a single error which is ascribed to the measure in which the sign originates.
5. Articulation (A)	a. A slur omitted, a staccato omitted, a slur carried onto notes which should be staccato, or a broken slur are all counted as errors.	(1) Occasionally one finds a student who does not interpret articulation when sight-reading. Do not mark articulation errors when you first discover this, but inform him/her it is his/her responsibility to distinguish between slurred and detached notes. After that, grade him as strictly as others. If in doubt, mark as error. The burden of proof lies with the student. See that he/she understands that and then score rigidly.
6. Rests (R)	a. Ignoring a rest or failure to give a rest its correct value is an error. Apply the same standards as to a sustained note (see rule 2).	
7. Holds and Pauses (R)	a. Holds written thus (fermata) should be treated in the same way as other expression marks (See section 4a, rule 1)	
	b. Pauses between measures, no matter how long, are not to be counted as errors; however, do not inform the student of this fact.	
	c. Pauses between notes within the measure are to be counted as errors.	
8. Pedaling (S)	a. A student is to be marked wrong if he fails to use the pedal in a piece where he is instructed to pedal.	(1) If a general instruction is given at the beginning of the piece to use pedaling which the student fails to observe, he is to be marked wrong only in the first measure of the piece.
		(2) If pedal markings appear in individual measures, each measure with an indication to pedal which is not pedaled is marked incorrect.
9. Hand Synchronization	a. A student is to be marked wrong if he does not play the notes in the RH with the appropriate notes in the LH.	

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.2. Comparing Assessment Methods

In order to compare the three assessment methods, the Piano Sight-reading Scale, sight-reading error quantifications and evaluation by expert examiners, it was necessary to have new sight-reading performances which would be assessed with the three methods. This section outlines our collection of sight-reading performances which would be used for the ensuing assessments by the three methods under comparison.

2.2.1. Participants

Eight piano students between the ages of 11 and 16, with the average being 14 years of age, were selected for this experiment (see Table 11).

Table 11. Participants Characteristics

Subject	Gender	Age	Level	No. of years of piano lessons
A	F	11	RCM 8	5
B	M	13	RCM 8	8
C	F	16	RCM 9	11
D	F	15	RCM 6	11
E	M	16	RCM 9	10
F	F	13	RCM 9	8
G	F	14	RCM 7	7
H	F	14	RCM 7	7

These subjects ranged from grades six to grade nine as defined by the Royal Conservatory of Music. The average grade level of the subjects was of 7.88. Participants had been playing the piano between five and 11 years, with an average of 8.25 years. Six of the participants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were girls and two were boys. They were all from the Ottawa-region. Piano teachers were first contacted and informed about this project. If interested, packages containing a letter of explanation as well as a participation consent form for both parent and student (see Appendixes C, D, E, for the letter of explanation and the two consent forms), were brought to them in order to be distributed to their students who fit the project's pre-determined criteria. Students who showed interest were then given a package containing the letter and consent forms and then were contacted by the project coordinator by phone or email to schedule a session date and time.

2.2.2. Procedure

The data collection took place at the Piano Pedagogy Research Laboratory at the University of Ottawa. This infrastructure, under the direction of Dr. Gilles Comeau, is a fully equipped piano studio with cameras, LCD screens, two Disklavier ProMark3 grand pianos all within a welcoming environment. The student, his/her parent and the test administrator were present during the experimentation. When the subjects entered the studio where the experiment was to take place, they were first asked to sign a consent form and were then instructed to fill out a short form about their age, number of years of instructions at the piano, the RCM exam level for which they were currently preparing, as well as a few questions relating to sight-reading (see Appendix F for student form). The students were informed that their performance on this test would not be seen by any one other than those involved in this research. Their participation was on a voluntary basis and they could discontinue their involvement at any time. Their parent was also asked to sign a consent form.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The subjects were instructed to sit at the piano, a Yamaha Disklavier 7'6" acoustic grand piano equipped with a MIDI operating system, as they would normally sit at the instrument when practicing or performing. They were then given the opportunity to adjust the bench in height and distance from the instrument depending on their needs. Once the subjects were comfortably seated, the project coordinator explained that they were going to sight-read pieces of increasing difficulty. They were told that they were required to sight-read the exercises as accurately as possible and to the best of their ability. A paper copy of the first piece, exercise 1, was then placed in front of them and they were asked to review the piece for 30 seconds. When the preview time was over they were instructed to start performing. Once they were done the first exercise, it was removed and a second piece, exercise 2, was placed in front of them. Again they were asked to review the piece for 30 seconds and then to begin playing. This process was repeated, and the pieces sight-read in order from exercise 1 to exercise 11, until the pieces were too difficult for the subject and all measures were being performed with errors. At this point, subjects were no longer given exercises to perform, but rather informed that the testing was over. The subjects were then congratulated for their effort and thanked for their involvement in the project.

The musical exercises were presented to participants on an 8 by 11 inch white sheet of paper just as music is frequently presented to participants on a sheet of paper in music reading research (Gilman, 2000; Gudmundsdottir, 2003) as well as in examination and competition settings. The students' performance was recorded as MIDI files on the Disklavier. The MIDI data was converted into written notation scores with the use of the software CuBase. These will be referred to as Cubase Musical Scores (CMS). Cubase, which generates a musical score that conforms to the performance, has previously been used

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in research settings (Gilman, 2000). With this software, wrong notes, rhythm miscues, and timing inconsistencies are visually apparent on the CMS. The Score Editor was used to organize the CMS for all pieces, under the Main Setting page it was indicated to not display "note overlap"; on the Polyphonic Setting page the staff mode was set to "split". Since pieces differed in terms of key signature, time signature, etc., other specifications were set independently for each piece (see Table 12).

Table 12. CuBase Specifications for Each Exercise

Exercise	Note and rest value	Key signature	Time signature
1	Eighth-note Eighth-rest	No alteration	No alteration
2	Eighth-note Eighth-rest	F major	¾ time
3	Eighth-note Eighth-rest	D major	No alteration
4	Eighth-note Eighth-rest	A major	No alteration
5	Eighth-note Eighth-rest	G major	¾ time

Each one of these settings clarified and organized the performance's musical score, simplifying the forthcoming quantification and made possible a more precise and consistent measurement of performances.

Audio recording was also done during participants' performance. The audio recordings were converted using MPEG Streamclip software to audio Compact Disks organized by subject. The audio files were also converted into Mp3 format to be sent via the internet also organized by subject. These audio recordings were used for the measurement with the Piano Sight-reading Scale as well as for the evaluation by expert examiners.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3

3. Data Presentation of Performance Assessment

Once the eight subjects' performances were collected, they were assessed with each of the three methods being studied: the Piano Sight-reading Scale (PSS), Gilman's scoring algorithm (GSA), and expert evaluation (EE). The following section describes the assessment of performances and presents the ensuing results and subject placements according to their final grades with each of the three methods.

To compare the three assessment methods it was decided to only retain the performances of the first five exercises of the PSS. The subjects who participated in the research project were of differing sight-reading abilities and therefore, attained various levels in the eleven pieces of our scale. Students were stopped anywhere between the fifth exercise and the eleventh and final exercise. Even though research using the WFPS usually followed its outlined performance measurement procedure (Elliott, 1982a, 1982b; Gaynor, 1995; Goolsby, 1994a; Gromko, 2004; McPherson, 1994), studies using error quantification and evaluation by expert examiners used a small number of exercises, between two and four, and always assessed all subjects on the same exercises.[4] It was therefore determined that measuring differing numbers of exercises with Gilman's Scoring Algorithm would impact the assessment of subjects. Furthermore, providing expert examiners with differing numbers of performed exercise for each subject would certainly have a significant affect on their

[4] Eaton (1978) analyzed the performance of four pieces to evaluate the sight-reading ability of his subjects. Gilman (2000) two pieces; Gudmundsdottir (2003) three; Killian and Henry (2005) two; Lehmann and Ericsson (1993) two. The only exception being Elliott (1982a), where all the exercises of the WFPS were used for the analysis. In research using error measurement and expert evaluations to grade performances, the same exercises were assessed for each subject (Gilman, 2000; Gilman, Underwood, and Morehen, 2002; Gudmundsdottir, 2003; Lehmann and Ericsson, 1993; Levy, 2001).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluation of sight-reading performances. Thus, it was decided that only the performance of the first five exercises of the PSS would be used. This way all subjects would be assessed on the same pieces with the three methods enabling an accurate comparison.

A pilot study was undertaken in order to verify the procedure of each method of assessment and to confirm the effectiveness of our comparison approach. For this pilot study two of the subjects' performances were assessed with each of the methods, PSS, GSA, and EE. For measurement with the PSS one graduate music student was given a print copy of the error classification document (see Table 8) as well as two print copies of each of the exercises one to five composed for the purpose of this project (see Appendixes G, H, I, J, and K for the exercises). The procedure was thoroughly explained to him and he was asked if he had any concerns or questions. Once the procedure, and error classification was clear he was provided with the audio recordings on a Compact Disk and asked to measure the performances. The two subjects were then placed in order of sight-reading performance according to their results with the PSS. Measurement with the GSA was done by the author of this thesis using Cubase Musical Scores. The quantification of pitch and rhythm errors were done separately. Subjects were then placed in order of performance according to their results with the GSA. Evaluation with EEs had one expert knowledgeable in sight-reading evaluation of the RCM assess the subjects' performances. This examiner had previously assisted in the revision of exercises composed for the purpose of the PSS. She was sent guidelines for the evaluation process along with sheets on which to evaluate the two subjects.

The pilot study revealed that it was possible to rate subjects with all three methods of assessment. No problems or concerns came out of the measurement with the PSS. However, some adjustments and additions needed to be made to the assessment procedures with the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

GSA and the EEs. After quantifying the performances with the GSA, it was felt that modifications were needed to account for variations in testing procedures. Gilman provided a tempo to pianists before they commenced their sight-reading performance while we did not. Since no tempo was provided to pianists in our experiment some undertook a very slow performance. In the music scores printed out with Cubase (CMS), slow performances were registered as having much longer note values, which in turn resulted in exceedingly low grades with the GSA. We felt that performances could not be penalized for a slow speed as no tempo had been given to subjects. This would be accounted for by adjusting the performance tempo in Cubase which resulted in CMSs displaying notes predominantly equivalent to their intended values in the original scores. The EE evaluation resulted in a few additions to the grading instructions. This examiner pointed out that no specifications were given as to what represented a passing and failing grade. She therefore used the RCM's grading designations. To remedy this omission, the following instruction was added to the package given to the EEs participating in the later evaluation: "For this project, we will go by the RCM standard and use 6/10 to indicate a passing grade. Any mark below 6 is a failing grade." The comments provided by this examiner on her evaluation sheet were thorough and informative and as she could not participate in the later evaluation, her comments served as examples for the three new examiners involved during the EEs portion of the performance assessment.

The pilot study confirmed that a comparison of the performance assessments with these three methods was possible. It also made apparent that a few adjustments and additions needed to be made in order to refine the measurement with the GSA and evaluation by EEs. The following three sections will now provide detailed accounts of the assessment of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

same performances with the Piano Sight-reading Scale, Gilman's Scoring Algorithm, and Expert Examiners. Sections also display the eight subjects' placement according to their final grades with the three methods of assessment.

3.1. Measurement According to the Piano Sight-reading Scale

Two doctoral students (Tester 1 and Tester 2), with Master of Arts, specializing in Piano Pedagogy, and Bachelors of Music, specializing in Piano Performance were recruited to measure all the performances using the Piano Sight-reading Scale (see Table 13 for the distribution of marks by both testers and the subject's average mark on each of the five exercises).[5] They were invited to review the PSS' guidelines and then asked if they had any questions or concerns, which were duly addressed by the researcher (see Appendixes L and M for letter of explanation and consent form given to graduate students). Once everything was made clear, the two testers were given the original musical scores of the five performances they were about to assess (see Appendixes G, H, I, J, and K for the musical scores). They were provided with one set of five musical scores for each of the eight subjects and instructed to assess the performances directly on the musical scores and then asked to quantify the results. They were instructed to grade performances measure by measure, identifying all errors made and classifying them according to the abbreviations provided on the PSS (i.e. (P) for pitch errors, (R) for time errors, etc.). Audio recordings of the sight-reading performances were then provided to the testers on Compact Discs labeled a1, a2, a3,

[5] Elliott (1982a) had two instrumental music education graduate students score subjects' taped performances with the WFPS. An average of the two scores was then used as each subject's final grade. Elliott, himself, then scored the sight-reading performances. McPherson (1994) also had two instrumental education graduate students score subjects' performances. Other researchers using the WFPS didn't elaborate on their choice of appraisers. In Elliott (1982a) and McPherson's (1994) research it was deemed necessary to have two evaluators in order to account for individual differences in the assessment process. Taking an average of both evaluators' results would ensure a more accurate implementation of the PSS.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

a4, a5 for the first subject, b1, b2, b3... for the second subject, and so forth for the subsequent subjects. Both testers received Compact Disks with subjects in differing orders. These orders were randomly organized by the researcher. This ensured that subjects would not receive undue assessments in relation to their placement on the Compact Disk.

Table 13. Measurement of Subjects with the Piano Sight-reading Scale

		Exercise 1	Exercise 2	Exercise 3	Exercise 4	Exercise 5	Total
Subject A	Tester 1	1	1	2	0	2	6
	Tester 2	1	1	1	0	2	5
	Average	1	1	1.5	0	2	5.5
Subject B	Tester 1	4	3	8	4	3	22
	Tester 2	4	3	6	2	4	19
	Average	4	3	7	3	3.5	20.5
Subject C	Tester 1	1	0	1	1	0	3
	Tester 2	2	1	1	0	0	4
	Average	1.5	0.5	1	0.5	0	3.5
Subject D	Tester 1	1	1	5	0	1	8
	Tester 2	1	0	4	1	2	8
	Average	1	0.5	4.5	0.5	1.5	8
Subject E	Tester 1	1	1	3	1	3	9
	Tester 2	1	1	2	1	2	7
	Average	1	1	2.5	1	2.5	8
Subject F	Tester 1	2	1	3	1	2	9
	Tester 2	2	1	3	1	1	8
	Average	2	1	3	1	1.5	8.5
Subject G	Tester 1	2	0	1	0	1	4
	Tester 2	1	0	0	0	1	2
	Average	1.5	0	0.5	0	1	3
Subject H	Tester 1	2	1	3	0	2	8
	Tester 2	2	1	2	0	2	7
	Average	2	1	2.5	0	2	7.5

Note. Exercise 1 was graded out of 4 marks; Exercise 2 out of 4 marks; Exercise 3 out of 8 marks; Exercise 4 out of 6 marks; Exercise 5 out of 8 marks; Total out of 30 marks

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The testers were informed that should questions arise during the grading process, they were to ask the researcher for clarification. The two testers' measurement of performances was then compared and an average of both grades was determined.

3.1.1. Measurement of Subjects

Subject A's performance of exercise one resulted in a grade of 1 out of 4 with both testers, this included pitch errors in two measures and an articulation in one measure. Exercise two was given a grade of 1 out of 4 by both testers, comprising rhythm errors in three measures and tester two also found articulation errors in the same three measures. Exercise three was given a grade of 2 out of 4 by tester one and 1 out of 4 by tester two for an average grade of 1.5, this included pitch errors in six measures, tester two also found a rhythm, and articulation error in one measure. Exercise four was given a grade of 0 out of 6 by both testers, this included five measures with pitch errors and one measure with rhythm errors. Exercise five was given a grade of 2 out of 8 by both testers, this included 6 measures with rhythm errors and one tester also had articulation errors.

Subject B's performance of exercise one was given a grade of 4 out of 4 by both testers. Exercise two was given a grade of 3 out of 4 by both testers, resulting from one measure with an articulation error. Exercise three was given a grade of 8 out of 8 by tester one and 6 out of 8 by tester two for an average of 7, tester two found rhythm errors in two measures. Exercise four was given a grade of 4 out of 6 by tester one and 2 out of 6 by tester two for an average of 3, resulting from one pitch error, one expression error and tester two also found rhythm errors in two measures. Exercise five was given a grade of 3 out of 8 by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tester one and 4 out of 8 by tester two for an average of 3.5, including four measures with articulation errors and one measure with a pitch error.

Subject C's performance of exercise one was given a grade of 1 out of 4 by tester one and 2 out of 4 by tester two for an average of 1.5, resulting from one measure with rhythm errors, one measure with pitch errors, and tester one also found one measure with an articulation error. Exercise two was given a grade of 0 out of 4 by tester one and 1 out of 4 by tester two for an average of 0.5, including three measures with pitch and rhythm errors, tester one also found one measure with rhythm errors. Exercise three was given a grade of 1 out of 8 by both testers, resulting from seven measures with pitch errors, three of which also included rhythm errors. Exercise four was given a grade of 1 out of 6 by tester one and 0 out of 6 by tester two for an average of 0.5, resulting from five measures with pitch errors, many of which also had rhythm errors, tester two also found expression errors and rhythm errors in the sixth measures. Exercise five was given a grade of 0 out of 8 by both testers, resulting from pitch, rhythm, articulation, and expression errors in all measures.

Subject D's performance of exercise one was given a grade of 1 out of 3 by both testers, including two measures with pitch errors and one measure with articulation errors. Exercise two was given a grade of 1 out of 4 by tester one and 0 out of 4 by tester two for an average of 0.5, resulting from rhythm and pitch errors in three measures, tester two also marked a rhythm error in the final measure. Exercise three was given a grade of 5 out of 8 by tester one and 4 out of 8 by tester two for an average of 4.5, this included rhythm and articulation errors in four measures. Exercise four was given a grade of 0 out of 6 by tester one and 1 out of 6 by tester two for an average of 0.5, including three measures with pitch errors and three measures with articulation errors. Exercise five was given a grade of 1 out

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of 8 by tester one and 2 out of 8 by tester two for an average of 1.5, this included pitch errors in three measures and rhythm errors in three measures, and tester one also found a timing error.

Subject E's performance of exercise one was given a grade of 1 out of 4 by both testers, including two measures with articulation errors and one with rhythm errors. Exercise two was given a grade of 1 out of 4 by both testers, including two measures with articulation errors and one with pitch errors. Exercise three was given a grade of 3 out of 8 by tester one and 2 out of 8 by tester two for an average of 2.5, including pitch, rhythm, articulation and timing errors. Exercise four was given a grade of 1 out of 6 by both testers, including three measures with pitch and rhythm errors, two with rhythm and expression errors. Exercise five was given a grade of 3 out of 8 by tester one and 2 out of 8 by tester two for an average of 2.5, including rhythm errors in five measures which also included pitch, articulation, and expression errors.

Subject F's performance of exercise one was given a grade of 2 out of 4 by both testers, including articulation errors in two measures. Exercise two was given a grade of 1 out of 4 by both testers, including three measures with articulation errors. Exercise three was given a grade of 3 out of 8 by both testers, including five measures with articulation errors. Exercise four was given a grade of 1 out of 6 by both testers, including two measures with pitch errors, four measures with articulation errors, and one with expression error. Exercise five was given a grade of 2 out of 8 by tester one and 1 out of 8 by tester two for an average of 1.5, including rhythm, expression, articulation, timing, and pitch errors.

Subject G's performance of exercise one was given a grade of 2 out of 4 by tester one and 1 out of 4 by tester two for an average of 1.5, including articulation errors. Exercise two

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was given a grade of 0 out of 4 by both testers, including one measures with an articulation error, two measures with rhythm errors, and one measure with a pitch error. Exercise three was given a grade of 1 out of 8 by tester one and 0 out of 8 by tester two for an average of 0.5, including rhythm errors in seven measures, pitch errors in three measures, and one articulation error. Exercise four was given a grade of 0 out of six by both testers, including rhythm errors in all measures, which also included articulation, expression, and pitch errors. Exercise five was given a grade of 1 out of 8 by both testers, including seven measures with rhythm errors, which also included articulation, pitch and expression errors.

Subject H's performance of exercise one was given a grade of 2 out of 4 by both testers, including two measures with articulation errors. Exercise two was given a 1 out of 4 by both testers, including three measures with articulation and rhythm errors. Exercise three was given a grade of 3 out of 8 by tester one and 2 out of 8 by tester two for an average of 2.5, including pitch errors in two measures and articulation errors in four measures. Exercise four was given a grade of 0 out of 6 by both testers, including pitch errors in four measures and rhythm errors in two measures. Exercise five was given a grade of 2 out of 8 by both testers, including a pitch error in one measure, rhythm errors in three measures, as well as timing and articulation errors.

3.1.2. Subject Placement

The disparities in subject measurement by the two testers were minimal; however, testers did at times find differing results in both final grades and types of errors. In three occasions, these differences between the two testers' final grades resulted in variations to the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

subjects' rankings (see Table 14 for subject rankings with both testers as well as their average ranking).

Table 14. Subject Placement with the Piano Sight-reading Scale

<table>
<tr><th>Ranking</th><th>Tester 1</th><th>Tester 2</th><th>Average</th></tr>
<tr><td>First place</td><td>B</td><td>B</td><td>B</td></tr>
<tr><td>Second place</td><td rowspan="2">E & F</td><td rowspan="2">D & F</td><td>F</td></tr>
<tr><td>Third place</td><td rowspan="2">D & E</td></tr>
<tr><td>Fourth place</td><td rowspan="2">H & D</td><td rowspan="2">E & H</td></tr>
<tr><td>Fifth place</td><td>H</td></tr>
<tr><td>Sixth place</td><td>A</td><td>A</td><td>A</td></tr>
<tr><td>Seventh place</td><td>G</td><td>C</td><td>C</td></tr>
<tr><td>Eighth place</td><td>C</td><td>G</td><td>G</td></tr>
</table>

Subject E tied for second and third place and subject D tied for fourth and fifth place with tester 1, these two subjects were inverted with tester 2's ranking. This resulted in subjects D and E tying for fourth and fifth place with the average ranking. Subject G was placed in the seventh rank and subject C was placed in the eighth rank by tester 1, this placement was inverted by tester 2 whose placement reversed tester 1's ranking. Nevertheless, these minimal disparities can be explained by individual differences resulting in unavoidable variances in musical backgrounds. For the comparison of the three methods of assessment, an average of the two testers' final grades was used. This average out of 30 was then calculated out of 100 in order to make possible a comparison of the three methods (see Table 15 for subject placement and their average grade out of 100 with the PSS).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 15. Average Subject Placement with the Piano Sight-reading Scale

Ranking	Subject	Grade
First place	B	68.33 %
Second place	F	28.33 %
Third place	D & E	26.67 %
Fourth place		
Fifth place	H	25.00 %
Sixth place	A	18.33 %
Seventh place	C	11.67 %
Eighth place	G	10.00 %

The assessment of performances with the PSS involved a straightforward and reasonably timed process, lasting approximately two hours. The two testers had a few inconsequential queries, all relating to uncertainties with the error classification, which were promptly addressed by the researcher. The most common types of error found with this measurement of performance were pitch, rhythm, and articulation errors. Expression, timing, and pause errors were also at times found. Since exercises 1 to 5 did not include pedal markings, no pedaling errors were found. In addition no hand synchronization, rests, or tempo errors were located. The minimal differences between the two testers' final grades did at times result a change in subject placement, nevertheless, an average of the two grades was used for the comparison of the three methods of assessment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.2. Quantification According to Gilman's Scoring Algorithm

The following section presents the procedure used for the quantification of performances as well as results obtained using Gilman's Scoring Algorithm (GSA). The Cubase Musical Scores (CMS) were used for this quantification process. Subjects' errors were quantified following the method proposed by Gilman (2000) (see Table 2). The researcher calculated the errors, which were categorized as pitch errors, and rhythm errors directly on the CMS. GSA also provides a category for added notes which are calculated in the rhythm grade. The CMS displayed the exact pitches performed as well as the exact time value of each note; accordingly there was little room for differences in grades. Furthermore, Gilman (2000) did not discuss having a second evaluator assess the CMS. Therefore, only one researcher undertook the assessment using the GSA. Two steps were required for this quantification process. First the researcher identified all pitch inaccuracies on the CMS; the distance between the pitch in the CMS and the intended note was then calculated. Each pitch was given one point, 1/12 of a point was deducted for each semi-tone between the pitch in the CMS and the pitch in the original score. The deducted points were then subtracted from the highest possible grade.[6] The researcher then assessed the rhythm grade by calculating the number of eighth note beats between the note in the CMS and the intended note in the original score. Each note was given one point, ½ of a point was deducted for each eighth note beat between the note in the CMS and the note in the original score. Added notes and rests were assessed by deducting from the rhythm grade ½ of a point for each eighth beat of these added notes and rests. As previously discussed in the pilot study, subjects were not penalized for their performance tempo, whether it was too fast or too slow. Therefore, the tempo of the performance was taken into consideration and with the use of Cubase, CMS'

[6] The highest possible grade was found by counting the number of pitches in the original score.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were disposed according to the tempo chosen by the subject. More specifically, CMS were either set to "half time" or "double time" which resulted in CMS that for the most part displayed notes in their intended length. Three CMS resulted in dotted-rhythms which made it impossible to set with the use of Cubase functions. These scores were transcribed by the researcher using the scheme used by Cubase. These adapted CMSs were used during our assessment with the GSA (see Table 16 for the distribution of marks resulting from the quantification of performances according to Gilman's Scoring Algorithm).

Table 16. Quantification of Subjects with Gilman's Scoring Algorithm

		Exercise 1	Exercise 2	Exercise 3	Exercise 4	Exercise 5	Total
Subject A	Pitch Errors	13.25	14	35.75	43.08	45.83	151.91
	Rhythm Errors	6.00	6,50	18.00	16.50	29.00	76.00
	Total	19.25	20,50	53.75	59.58	74.83	**227.91**
Subject B	Pitch Errors	14.00	14.00	37.00	44.00	46.58	155.58
	Rhythm Errors	11.00	7.50	23.00	37.00	37.50	116.00
	Total	25.00	21.50	60.00	81.00	84.08	**271.58**
Subject C	Pitch Errors	12.33	12.33	35.08	40.5	45.83	147.07
	Rhythm Errors	4.50	6.50	-48.00	14.50	-74.50	-97.00
	Total	16.83	18.83	-12.92	55.00	-28.67	**49.07**
Subject D	Pitch Errors	7.00	5.92	37.00	43.92	46.00	139.84
	Rhythm Errors	9.50	7.50	10.50	13.50	32.00	73.00
	Total	16.50	13.42	47.50	57.42	78.00	**212.84**
Subject E	Pitch Errors	14.00	13.83	37.00	44.00	47.00	155.83

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

	Rhythm Errors	6.00	10.00	1.00	2.00	32.50	51.50
	Total	20.00	23.83	38.00	46.00	79.50	**207.33**
Subject F	Pitch Errors	14.00	14.00	37.00	44.00	47.00	156.00
	Rhythm Errors	11.00	10.00	20.00	36.00	-4.50	72.50
	Total	25.00	24.00	57.00	80.00	42.50	**228.50**
Subject G	Pitch Errors	14.00	13.00	35.92	44.00	46.58	153.5
	Rhythm Errors	9.50	9.00	-4.00	-10.00	-0.50	4.00
	Total	23.50	22,00	31.92	34.00	46.08	**157.50**
Subject H	Pitch Errors	14.00	14.00	37.00	42.99	46.83	154.82
	Rhythm Errors	13.00	8.50	21.00	-20.50	36.50	58.50
	Total	27.00	22.50	58.00	22.49	83.33	**213.32**

Note. Exercise 1: Pitch errors calculated out of 14; rhythm errors calculated out of 14; total points calculated out of 28; Exercise 2: Pitch errors calculated out of 14; rhythm errors calculated out of 14; total points calculated out of 28; Exercise 3: Pitch errors calculated out of 37; rhythm errors calculated out of 37; total points calculated out of 74; Exercise 4: Pitch errors calculated out of 44; rhythm errors calculated out of 45; total points calculated out of 89; Exercise 5: Pitch errors calculated out of 47; rhythm errors calculated out of 55; total points calculated out of 102

3.2.1. Quantification of Subjects

Pitch Errors

Pitch errors were first identified for each exercise. These were then calculated by measuring the number of semi-tones between the pitch indicated in the original grade and the pitch in the CMS, 1/12 of a point is deduced for each semi-tone.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject A's performance of exercise one included five pitch errors which reduced the grade by 9/12 of a point. Exercise two had no pitch errors. Exercise three had 7 pitch errors which reduced the final grade by 15/12 of a point. Exercise four had 9 pitch errors which reduced the grade by 9/12 of a point. And exercise five had 1 pitch errors as well as one missing note which reduced the grade by 1 and 10/12 of a point.

Subject B's performance of exercise one through four included no pitch errors. The performance of exercise five included 2 pitch errors which reduced the grade by 5/12 of a point.

Subject C's performance of exercise one included 2 pitch errors which reduced the grade by 8/12 of a point. Exercise two included six pitch errors which reduced the grade by 20/12 of a point. Exercise three included 10 pitch errors as well as one missing note which reduced the grade by 1 and 11/12 of a point. Exercise four included 10 pitch errors as well as two missing notes which reduced the grade by 2 and 18/12 of a point. And exercise five included two pitch errors and one missing note which reduced the grade by 10/12 of a point.

Subject D's performance of exercises one contained 7 pitch errors which reduced the final grade by 7 points. Exercise two contained 8 pitch errors which reduced the grade by 7 points and 10/12 of a point. Exercise three contained no pitch errors. Exercise four contained 1 pitch errors which reduced the grade by 1/12 of a point. And exercise five had one missing note which reduced the grade by 1 point.

Subject E's performance of exercise one contained no pitch errors. Exercise two contained 2 pitch errors which reduced the grade by 2/12 of a point. And exercises three, four, and five contained no errors.

Subject F's performance of exercises one through five contained no pitch errors.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject G's performance of exercises one contained no pitch errors. The performance of exercise two was missing a note which reduced the grade by 1 point. Exercise three contained 1 pitch error and one missing note which reduced the grade by 1 and 1/12 of a point. Exercise four contained no pitch errors. And exercise five contained 2 pitch errors which reduced the grade by 5/12 of a point.

Subject H's performance of exercise one through three included no pitch errors. Exercise four contained 5 pitch errors which reduced the grade by 1 point. Exercise five contained 2 pitch errors which reduced the grade by 2/12 of a point.

Rhythm Errors

The quantification of pitch errors was followed by a calculation of the rhythm and added note errors. Rhythm errors were calculated by counting the number of eighth-note beats between the performed note found in the CMS and the note indicated in the original grade. For each eighth-note beat added or subtracted from a note or rest during the performance, ½ of a point was deducted from the final grade. Added notes or rests were quantified by calculating the number of eighth-note beats in these added notes or rests, ½ of a beat was subtracted from the final grade for each eighth-note beat of an added note or rest.

Subject A's performance of exercise one included 2 added eighth-rests which reduced the grade by 1 point, and 9 notes were lengthened by a eighth-note beat which reduced the final grade by 4.5, one note was lengthened by a quarter-note beat and another by a dotted quarter-note beat which reduced the final grade by 2.5 points for a total subtraction of 8 points. Exercise two included 4 added eighth-rests which reduced the grade by 2 points, all eighth notes in the musical grade were performed as quarter-notes and the final note of the piece was held for a quarter-beat longer than its intended length which reduced the final

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

grade by 7.5 points. Exercise three contained many added rests, two added notes, and a few notes were held slightly longer than their intended length which reduced the final grade by 19 points. Exercise four contained many added rests and a few notes were held longer than their intended note value and few were not held long enough which reduced the final grade by 28.5 points. Exercise five contained many added rests and some notes were held slightly longer than their intended length which reduced the grade by 26 points.

Subject B's performance of exercise one included 1 added eighth-rests which reduced the grade by 0.5 points and 5 notes were lengthened by an eighth-note which reduced the final grade by 2.5 points for a total grade reduction of 3 points. Exercise two included 2 added eighth-rests and 2 added quarter-rests which reduced the grade by 3 points, some of the notes were held longer than their intended length which reduced the grade by 6.5 points. Exercise three contained a few added rests and some notes were held slightly longer than their intended length which reduced the grade by 14.5 points. Exercise four contained a few added eighth-rests and one added note, a few notes were held slightly longer than their intended value which reduced the grade by 8 points. Exercise five contained a few added rests and a few notes were held longer than their intended length and a few were not held long enough which reduced the grade by 17.5 points.

Subject C's performance of exercise one contained added notes and rests, a few notes were held longer than their intended length which reduced the grade by 9.5 points. Exercise two included many added rests and a few notes were held longer than their intended length which reduced the grade by 7.5 points. Exercise three contained many added notes and rests, including a false start, many notes were also held longer than their intended length which reduced the grade by 85 points. Exercise four also contained many added notes and rests a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

few notes were also held slightly longer than their intended length which reduced the grade by 30.5 points. Exercise five contained many added notes and rests, including a false start, many notes were also held longer than their intended length which reduced the grade by 129.5 points.

Subject D's performance of exercise one contained a few added rests, and two notes were held slightly longer than their intended length which reduced the grade by 4.5 points. Exercise two included four added eighth-rests, one added quarter-rest, three added notes, a few notes were held longer than their intended length and a few notes were not held long enough which reduced the grade by 6.5 points. Exercise three contained many added rests, many notes were also held longer than their intended length which reduced the grade by 26.5 points. Exercise four contained many added rests, a few added notes, and a few notes were not held long enough which reduced the grade by 31.5 points. Exercise five contained a few added notes and rests, and a few notes were held longer than their intended length which reduced the final grade by 23 points.

Subject E's performance of exercise one contained two added quarter-rests and three notes were held longer than their intended length which reduced the grade by 8 points. Exercise two included two quarter-rests, one eighth-rest, and a few notes were not held long enough which reduced the grade by 4 points. Exercise three contained many added rests, and a few added notes; a few notes were performed faster than their intended length which reduced the grade by 36 points. Exercise four contained many added notes and rests, some notes were held slightly longer and some performed longer than their intended length which reduced the grade by 43 points. Exercise five contained many added rests; a few notes were held longer than their intended length which reduced the grade by 22.5 points.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject F's performance of exercise one contained one added eighth-rest and a few notes which were held slightly longer than their intended length which reduced the final grade by 3 points. Exercise two included three added eighth-rests and a few notes were held longer than their intended length which reduced the grade by 4 points. Exercise three contained a few added rests and a few notes were held longer than their intended length and a few were not held long enough which reduced the grade by 17 points. Exercise four contained two added notes and a few rests and a few notes were held longer than their intended length and a few were not held long enough which reduced the grade by 9 points. Exercise five contained a few added notes and rests, including a false start; a few notes were held slightly longer than their intended length and a few notes were not held long enough which reduced the grade by 59.5 points.

Subject G's performance of exercise one contained three added eighth-rests and five notes were held slightly longer than their intended length which reduced the grade by 4.5 points. Exercise two included one added rest and a few notes were held slightly longer than their intended length and a few were not held long enough which reduced the grade by 5 points. Exercise three contained one added note and a few added rests, some notes were held longer than their intended length which reduced the final grade by 41 points. Exercise four contained one added note and many added rests; many notes were held longer than their intended value which reduced the grade by 55 points. Exercise five contained many added rests, again many notes were held longer than their intended value reducing the final grade by 55.5 points.

Subject H's performance of exercise two had notes which were held slightly longer than their intended value which reduced the grade by 1 point. Exercise two contained a one

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

added eighth-rest, and a few notes were held longer than their intended length which reduced the grade by 5.5 points. Exercise three contained a few added notes and rests; a few notes were also held longer than their intended length and a few were not held long enough which reduced the grade by 16 points. Exercise four contained a few added notes and rests; a few notes were held slightly longer than their intended length which reduced the grade by 67.5 points. Exercise five contained many added notes and rests and many notes were held longer than their intended length which reduced the grade by 18.5 points.

3.2.2. Subject Placement

In order to compare grades with the other two methods of assessment, final grades out of 321 were calculated out of 100 as a common denominator. Subjects were then placed in order of sight-reading performance according to their final grade with the GSA (See Table 17 for subject ranking and final grades out of 100 with the GSA).

Table 17. Subject Placement with Gilman's Scoring Algorithm

Ranking	Subject	Score
First place	B	84.60 %
Second place	F	71.18 %
Third place	A	71.00 %
Fourth place	E	64.59 %
Fifth place	H	66.45 %
Sixth place	D	66.38 %
Seventh place	G	49.07 %
Eighth place	C	15.29 %

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The quantification of pitch errors and rhythm errors with the use of the GSA was an extremely lengthy and complicated process (see Appendixes N to R for one example of quantification with the use of the GSA for each of the five exercises). Subjects lost inconsequential marks for pitch inaccuracies. Interestingly, key signature errors and lack of observation of accidentals were the types of pitch errors resulting in the least amount of mark deductions. Among the assessments, performance at the wrong octave resulted in the highest mark deduction by far. In the rhythm grade, long pauses and added notes and rests resulted in important mark deductions, whereas the performance of inaccurate rhythms (i.e. failing to perform eighth-notes twice as fast as quarter-note) did not necessarily result in low marks.

3.3. Evaluation by Expert Examiners

This evaluation required the expertise of three piano examiners from the Royal Conservatory of Music of Toronto (see Table 18 for the three Expert Examiners' marking of subjects). Two of these were senior examiners (Examiner 1 and Examiner 2) and one was an examiner in training at the Royal Conservatory of Music (Examiner 3). All examiners are well respected piano teachers with over 25 years of experience each. Examiners were contacted by email or telephone and informed about this project. Once they showed interest, they were either delivered packages or sent emails containing a consent form (see Appendix N for the Examiner Consent Form), a document containing instructions for the evaluation process, eight sheets on which to evaluate subjects, as well as one sheet on which to place subjects in order from best to weakest sight-reading performance (see Appendix O for the Examiner Scoring Sheets). These packages also included a Compact Disk with audio clips.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When the documentation was sent by email, the packages included MP3 audio files. Examiners were instructed that the RCM standard would be used to indicate a passing grade[7], and according to RCM grading procedures, grades could have half points.

Table 18. Evaluation of Subjects by Expert Examiners

		Exercise 1	Exercise 2	Exercise 3	Exercise 4	Exercise 5	Total
Subject A	EE 1	8.5	8	7.5	6.5	6	36.5
	EE 2	6.5	5.5	6	4	4	26
	EE 3	8	6	7	5.5	6	32.5
Subject B	EE 1	10	10	9.5	9.5	9	48
	EE 2	10	9.5	10	7.5	7.5	44.5
	EE 3	10	10	10	9.5	8.5	48
Subject C	EE 1	9	7.5	5.5	5	6	33
	EE 2	6.5	3	3	4	3	19.5
	EE 3	8	6.5	5.5	5.5	5.5	31
Subject D	EE 1	9.5	6.5	9	8	8	41
	EE 2	6.5	4.5	7.5	6.5	5.5	30.5
	EE 3	9	6.5	8.5	7	6.5	37.5
Subject E	EE 1	9.5	8.5	8	7.5	7	40.5
	EE 2	9.5	8	7.5	5.5	6.5	37
	EE 3	8.5	7	7.5	7	6.5	36.5
Subject F	EE 1	10	9.5	9.5	9.5	9	47.5
	EE 2	10	10	9.5	8.5	8	46
	EE 3	10	10	10	9	8.5	47.5
Subject G	EE 1	10	8.5	7.5	6	6.5	38.5
	EE 2	10	5.5	5	4	4	28.5
	EE 3	10	7	7	6.5	6	36.5
Subject H	EE 1	10	9	9	8	8.5	44.5
	EE 2	10	7	6.5	6	8.5	38
	EE 3	10	7.5	8.5	7	8	41

Note. All exercises were graded out of 10; Final scores were graded out of 50.

[7] A student requires a grade above 60% to pass. Any mark below 60% is a failing grade.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

They were then instructed to listen to each audio clip provided and grade that particular performance out of ten. Examiners were asked to comment on their grading strategies as well as their reasoning behind each awarded grade. Once they had listened to all clips from each subject, examiners were instructed to calculate the subject's final grade for a total out of fifty. These instructions were accompanied by examples of comments and reasoning on grading strategies provided from the experiment's pilot study with a different senior examiner from the RCM.

3.3.1. Subject Placement

After listening to all audio clips, examiners were asked to rate subjects according to their overall performance on the five exercises and to place them in order from best to weakest sight-reading performance. In order to compare final grades determined by EEs with final grades resulting from the other two methods of assessment, grades out of 50 were calculated out of 100 (see Table 19 for subject placement with Expert Examiners and final grades out of 100).

Interestingly, there is an important difference between the grading of EE 2 and that of both EE 1 and 3 (i.e. Subject G was given a final grade of 77% by EE1 and 73% by EE3, while EE2 awarded this same subject a 57%, a difference of 20% with EE1 and 16% with EE3). Not only does this stricter grading impact the subjects' final grades, it also results in variances to subject placement. Interestingly, both EE1 and EE3 had almost identical subject placements with one exception, EE1 ranked subject E in fifth place and subject G in sixth place while these two subjects tied for fifth and sixth place with EE3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 19. Subject Placement with Expert Examiners

	Examiner 1		Examiner 2		Examiner 3	
Ranking	**Subject**	**Score**	**Subject**	**Score**	**Subject**	**Score**
First place	B	96 %	F	92 %	B	96 %
Second place	F	95 %	B	89 %	F	95 %
Third place	H	89 %	H	76 %	H	82 %
Fourth place	D	82 %	E	74 %	D	75 %
Fifth place	E	81 %	D	61 %	E & G	73 %
Sixth place	G	77 %	G	57 %		
Seventh place	A	73 %	A	52 %	A	65 %
Eighth place	C	66 %	C	39 %	C	62 %

In contrast, EE2's placement of subjects varied from the other two examiners in four instances, in the first, second, fourth, and fifth place. Interestingly, at times this one EE's varying placement altered the other two EEs equivalent ranking (i.e. both EEs 1 and 3 ranked subject B in first place and subject F in second place EE 2's grading of these two subjects altered their position; subjects D and E, where both EEs 1 and 3 ranked subject D in fourth place and subject E in fifth place, and again EE 2's grading of these subjects altered their position) (see Table 20). For this project, an average of all three grades will be used for our final mark with EE evaluation in order to follow grading procedures used in other studies using more than one evaluator (Elliott, 1982a) (see Table 20 for the average subject placement and average final grades out of 100 for EEs).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 20. Average Subject Placement with Expert Examiners

Ranking	Subject	Score
First place	F	94 %
Second place	B	93.66 %
Third place	H	82.34 %
Fourth place	E	76 %
Fifth place	D	72.66 %
Sixth place	G	69 %
Seventh place	A	63.33 %
Eighth place	C	55.66 %

The evaluation of sight-reading performances by EEs was a straightforward and fairly rapid process; however, it did involve complex intellectual assessment of performances not necessary for the other two methods, which were accompanied by pre-existing error measurement and quantification guidelines. The evaluation by EEs was done with audio recordings as this corresponded to several studies using this method of sight-reading performance assessment (Eaton, 1978; Gilman, 2000; Lehmann & Ericsson, 1993; Rogers, 1996; Salis, 1977). All three EEs received Compact Disks or Mp3 files via the internet with subjects in differing orders. The orders were randomly organized by the researcher. This ensured that subjects' placement on the Compact Disk or as Mp3 files would not affect their final grade. Interestingly, many variances were found among the final grades determined by EEs 1 and 3 and those determined by EE2, at times these altered the final placement of subjects. EEs for the RCM undergo rigorous training in order to provide reliable and standardized judging among evaluators. However, in many examination and educational settings, EEs would typically assess subjects in person without a recording medium. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

setting used in this study was most likely less familiar for EEs, and therefore, might have had an impact on their grading.

3.4. Conclusion

After a pilot study, a detailed assessment of the sight-reading performances was undertaken with each method, PSS, GSA, and EE. The measurement of performances with the PSS by two graduate students was a fairly rapid and straightforward process resulting in minimal variations between the two testers. An average of both testers' final grades was calculated in order to be used for the comparison of the three methods of assessment. Quantification of performances with the GSA by this researcher was a lengthy and complicated process which was slightly modified after our pilot study. Nevertheless performances were measured in relation to pitch and rhythm errors and final grades were determined. Evaluation by three EEs from the RCM was a rapid process; interestingly, there were significant disparities in the final grades awarded by EEs. Nevertheless, in standing with previous research, an average of the three grades was calculated to be used for the comparison of methods. The next chapter will relay our comparison of the three methods of assessment in relation to their grading procedures and their assessment of piano sight-reading performances.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4

4. Data Comparison

The purpose of this study is to compare three methods of sight-reading assessment used in music reading research, the WFPS which was adapted into the PSS, the GSA, and expert examiners, to determine if they provide analogous grading of piano sight-reading performance. In order to do so, two steps were required: 1) a comparison of assessment procedures used with each method; and 2) a comparison of subject assessment obtained by each method.

4.1. Comparison of Assessment Procedures

The three methods of assessment being compared in this research had very different procedures and strategies for grading subjects. Three main distinctions are apparent: 1) the type of error being considered; 2) the allocation of marks; and 3) the length and complication of carrying out an assessment of performance with each method.

4.1.1. Types of Errors

The types of errors being considered differed significantly in the three methods of assessment. The PSS took into consideration the many various types of errors which could be made during sight-reading performance at the piano. Its error classification provided a thorough and detailed description of these errors, which included pitch, rhythm, timing, expression, articulation, rests, holds and pauses, pedaling, and hand synchronization. Interestingly, the predominant types of errors recorded by the two testers were pitch errors,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

articulation errors, timing errors, and rhythm errors. Expression errors were also at times noted. No hand synchronization or tempo errors were noted, and no rest, and hold errors were found. The five exercises used as stimuli did not contain pedal markings; therefore, no pedaling errors were noted. On the other hand, the GSA only took into consideration pitch errors and rhythmic and timing inaccuracies, which included added notes, rests, and pauses, and did not measure or even discuss elements such as articulation, expression, hand synchronization, etc. The GSA stressed pitch and rhythm errors providing a very specific and complex algorithm for their measurement, but other elements, usually referred to as expressive elements (i.e. dynamics, articulation, etc.), were never considered. EEs fell somewhere in between the PSS and the GSA taking into account more types of errors than the GSA but not as systematic as the PSS. EEs almost always mentioned pitch, rhythmic, and timing inaccuracies, they also often noticed articulation and expression errors as well as issues with the performance tempo. They also at times noted pauses and added notes. Though these were not always discussed, the types of errors analyzed in the evaluation by EEs resembled more closely the PSS than the GSA.

4.1.2. Allocation of Marks

The allocation of marks also varied according to grading strategies used by the three methods of assessment. The PSS did not differentiate between the different types of errors providing an equal subtraction of marks regardless of the types of errors made. With the PSS a pitch or rhythm error was granted the same weight as an expression or an articulation error. This grading procedure makes possible the allocation of zero as a final mark resulting solely from expression or articulation errors. With the PSS, performances were graded by measure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This resulted in performances being deducted uneven marks for the same number of errors, as well as equal penalization despite the number of errors performed. For example, a performance containing one error in each measure resulted in a grade of zero, whereas, a performance containing the same number of errors within only a few measures resulted in a grade merely reduced by a few points. In contrast, measurement according to the GSA, even though it only took into consideration pitch and timing inaccuracies, did distribute mark subtractions differently depending on the types of errors performed and measured performances note by note. However, the GSA's distribution of marks resulted in minimal grade reductions resulting from pitch errors. With this algorithm, if an exercise was performed entirely on wrong pitches, the pitch grade could still be relatively high if the interval between the performed pitch and the pitch indicated in the original score was reasonably small. On the other hand, measurement with the GSA resulted in particularly low rhythm grades. These were especially low when the performance included numerous added notes, rests and pauses; such errors could at times result in rhythm grades below zero. Interestingly, subjects could receive minimal grade reductions resulting from rhythm errors if the tempo was steady and the performance did not include lengthy pauses or added notes. Evaluation by EEs was a more subjective process resulting in more indefinite allocations of marks than the PSS and the GSA since EEs were not provided with criteria or set guidelines for the evaluation of performances. Nevertheless, they did stress rhythm errors as well as pitch errors associated to the key signature. They also at times noted articulation errors, expression errors, as well as tempo errors, among other.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.1.3. Length and Complexity

The grading procedures of the PSS, the GSA, and EEs also differed in their complexity and the amount of time necessary for their undertaking. The PSS resulted in a reasonably straightforward and rapid process. The instructions were clear and the testers had little difficulty understanding the error classification. The measure as grading unit generated a very easy and uncomplicated calculation of marks. The testers, therefore, were not required to assume complex mathematical equations for the accumulation of grades. The GSA, on the other hand, resulted in an intricate and lengthy procedure. Even though the error classification for this assessment method was straightforward, the mathematical equations required for the calculation of marks complicated this method. These equations necessitated intricate calculations for the pitch grades where the evaluator had to measure each semi-tone between the pitch in the CMS and the pitch in the original score providing a fraction on 12. For example, if the subject performed a C instead of a C#, the pitch grade was reduced by 1/12 of a point, if the subject performance a C instead of a D, the pitch grade was reduced by 2/12 of a point. These deductions were then accumulated to provide the pitch grade. A similar mathematical formula was used to calculate the rhythm score, however, rhythm errors were measured by ½ point which facilitated the grading with the GSA. Nonetheless, this algorithm resulted in a time consuming and intricate grading process. Evaluation by EEs, seeing as this method was not as controlled as the PSS or the GSA, resulted in a relatively fast and uncomplicated assessment. However, this method required an intricate thought process and the assignment of a grade for each performance which necessarily added to its time consumption.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.1.4. Conclusion

The difference in procedures and strategies required for the grading of subjects with the PSS, the GSA, and the EEs is apparent when looking at the types of errors being considered by each method, their allocation of marks, and the difficulty and complication of carrying out an assessment with each method. The PSS and the EEs have more thorough error classifications taking all aspects of performance into consideration when assessing subjects, whereas the GSA provides a more limited classification of errors to be measured. The allocation of marks by both the PSS and the GSA differs greatly as the first makes no distinction between different types of errors and uses the measure as grading unit, whereas the second provides a very specific and regulated differentiation among errors and uses individual notes as grading unit. Evaluation by EEs, though not as definite, did seem to stress pitch and rhythm errors over other types of errors such as articulation and expression. Finally, the complexity and length of the grading procedure further distinguishes these methods, the PSS and EEs assessments being relatively straightforward and rapid, whereas the GSA was a complex and time consuming process.

4.2. Comparison of Subject Assessment

The present section compares the assessment of subjects as obtained by the three methods. This comparison required three steps: 1) the comparison of grade deductions in relation to the types of errors found by each method; 2) the comparison of final grades determined by each method; and 3) the comparison of subject placement with each of the three methods.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.2.1. Grade Deductions

This section consists of an analysis of the three methods' assessments in relation to the types of errors found and the grade reductions resulting from these errors (see Appendixes P to W, for charts with exact grade deductions for each subject). This section will present specific information relating to the assessment of performances organized in three sections: 1) pitch errors; 2) rhythm errors; and 3) other errors. Even though it is not possible to provide exact causes of grade reductions by EEs, through their comments informed deductions were made and are presented in the following section.

Pitch Errors

Pitch errors were among the predominant types of errors found within the eight subjects' performances. All three methods of assessment found numerous pitch errors, nevertheless, the amount of point deductions for these errors were very different with each of the three methods. Through the comparison of grade deductions, it was found that the PSS, at times, subtracted much larger amounts than both the GSA and EEs. The largest discrepancies occurred between the assessment of the PSS and the GSA. One such instance occurred with subject C's performance of exercise four, where pitch errors were the only types of errors found in five out of the six measures by the PSS, reducing the final grade by 83.33%[8], whereas with the GSA, pitch errors reduced the final score by 3.5 out of 89 points, or 3.93% of the total grade, a difference of almost 80%. Furthermore, in a few cases, when pitch errors were found in the assessment with the PSS and were noted by EEs, these did not appear in the assessment with the GSA. One such instance occurred in the assessment of subject E's performance of exercise four, where half of the measures were found to contain

[8] In order to compare grade deductions it was necessary to place these numbers on a common denominator. This was done by finding percentages for these amounts in relation to the highest possible grade. The following equation was used: grade deduction ÷ highest possible grade x 100. (i.e. 5 ÷ 6 x 100 = 83.33%)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

solely pitch errors with the PSS, reducing the final grade by 50%, and all EEs noted pitch errors, when no such errors were found with the GSA. This resulted from the inclusion of added note errors within the rhythm error category with the GSA, creating further disparities in the comparison of pitch errors found with the three methods.

An additional pitch error discrepancy was observed between the assessment with the GSA and EEs. With the GSA, if a pitch error is very close to the intended note (i.e. F-F#, B-Bb), the grade reduction is minimal, therefore, when pitch errors resulted from a lack of observation to the key signature, few marks were lost. On the other hand, EEs in their comments stressed this type of pitch error above all others. An example of this can be seen with the assessment of subject D's performance of exercise four, where with the GSA, 0.09% was deducted from the total grade for pitch errors, while EEs, on average, reduced the final grade by 28.33%, stressing the subjects' problems with the key signature. Through this comparison of reductions resulting from pitch errors it was observed that the PSS often deduced more important marks than both the GSA and EEs, and that the GSA and EEs placed differing degrees of importance to pitch errors resulting from the key signature.

Rhythm Errors

Rhythm errors were also among the leading types of errors found within the eight subjects' performances. All three methods of assessment found a large number of performances to have rhythm inconsistencies, and although these differences were not as significant as those observed with pitch errors, again the amounts deducted from the performances' final grades varied with the three methods. A few instances occurred where the PSS deduced significantly more grades than both the GSA and EEs. One example of this occurred with subject A's performance of exercise two, where the PSS reduced the grade by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

75% for rhythm inaccuracies, whereas the GSA and EEs only diminished the final grade by 26.79% and 35% for the same errors. However, there were also a few instances where grade deductions were higher with the GSA than with the PSS and EEs. For instance, an extreme example occurred in the assessment of subject C's performance of exercise three, where the GSA subtracted 114.86% from the final grade resulting from rhythm errors, while only 12.5% was subtracted with the PSS' and 53.33% with EEs, however, resulting from pitch, rhythm and timing errors. Nevertheless, much fewer instances occurred where grade reductions resulting from rhythm errors varied as significantly as with pitch errors. It was also observed that the GSA consistently found performances to contain rhythm errors even, when both the PSS and EEs found no such errors. One example occurred in subject B's performance of exercise one, where both the PSS and EEs found the performance to contain no errors, while the GSA subtracted 10.71% from the total grade for rhythm errors.

An additional observation to the assessments in relation to rhythm errors was found between the quantification with the GSA and the evaluation by EEs. The GSA graded performances most severely for pauses, hesitations, and added notes and rests. However, in a few occurrences where subjects performed without taking into consideration the rhythm, performing all notes at the same length whether they were half-notes, quarter-notes, or eighth-notes, subjects did not necessarily receive very low marks if the performance did not contain pauses, hesitations, or added notes. On the other hand, EEs in their comments stressed this type of rhythmic error and graded these performances more strictly. One such example occurred in subject A's performance of exercise two, where the GSA only deduced 26.79% from the final grade for rhythm errors, while the PSS deduced 75% and EEs 35%, mainly stressing rhythm in their comments. Through this comparison of deductions resulting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from rhythm errors it was observed that even though extreme differences did not occur as often as with pitch errors, some were found where the PSS graded most severely and others where the GSA graded most severely. It was also found that even though the GSA consistently faulted performances' rhythm, the type of rhythm errors stressed were not necessarily equivalent to those stressed by EEs.

Other Errors

The GSA only measured pitch and rhythm, therefore, no other types of errors were found. But, the PSS and EEs found performances to contain other types of errors in addition to pitch and rhythm. The most predominant types of other errors found by the PSS and EEs were articulation, expression, and timing, however, timing can easily be linked to rhythm errors. Through the examination of such errors, many inconsistencies were found between the methods' assessments of performances. Firstly, these types of errors were often found in only one of the methods, while the same error was not observed by the other methods. One such example occurred in subject D's performance of exercise four where the PSS subtracted 41.67% for articulation errors, while no other method mentioned this type of errors; and EEs deduced 28.33% from the total grade, in part, because of expression errors, while these were not mentioned by the other methods. It was also observed that the PSS often deducted extreme marks resulting from articulation or expression errors, while these were rarely stressed by EEs, and not considered during the measurement with the GSA. One example occurred in subject F's performance of exercise two, where 3 out of the exercise's 4 measures were found to exclusively contain articulation errors, reducing the total grade by 75%. In this performance, EEs did mention articulation errors; however, only 16.67% was deduced from the total grade because of them. Through this comparison of other errors, it

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was observed that these types of errors are often found within one method's assessment when not in the others. Furthermore, the PSS often deduces many marks from the total grade for such errors when these are rarely emphasized by EEs, and not considered as types of errors by the GSA.

Conclusion

This comparison of each method's analysis of pitch errors, rhythm errors, and other errors, put forth many observations displaying the disparities between these three methods' assessments of the same performances. Pitch errors were found to be consistently graded most severely by the PSS, deducting significantly more marks than both the GSA and EEs. The GSA and EEs were observed to grade key signature errors very differently, the GSA grading these pitch errors most leniently, whereas EEs emphasized these as very important errors. Rhythm errors did not differ as widely as pitch errors, however, at times the PSS graded most severely and in other instances the GSA graded most severely. Interestingly, the GSA consistently found performances to contain rhythm errors, however, did not necessarily stress the same type of rhythm errors as EEs. Finally, articulation and expression errors were found to be the predominant type of other error, however, assessments were often found to differ in the types of errors found, and at times the PSS deduced extreme percentage of marks for such errors when these were not stressed by EEs and not considered by the GSA.

4.2.2. Final Grades

The following section compares the final grades obtained by each of the three methods. This comparison of final grades will be organized in two sections: 1) observations of the final grades from the highest to the lowest; and 2) observations between the three

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

methods' final grades in relation to each subject. The scoring unit differed with each method of assessment and, therefore, the total grades varied significantly (see Table 21 for grading units according to each method of assessment).

Table 21. Grading Units According to Each Method of Assessment

	PSS	GSA	Expert Evaluation
Exercise 1	4	28	10
Exercise 2	4	28	10
Exercise 3	8	74	10
Exercise 4	6	89	10
Exercise 5	8	102	10
Total	30	321	50

To compare the final grades obtained with each method of assessment, it was necessary to place the final grades on a common denominator. Therefore, all final grades were calculated out of 100 in order to analyze the results in terms of percentages (see Table 22 for subject's final grades). For this comparison it is important to keep in mind that the three methods were not intended for the same purposes and this necessarily had an impact on the final grades.[9]

Through the examination of final grades from the highest to the lowest many disparities were observed, however, some consistency was also noted. Firstly, two noteworthy hiatuses occurred among the final grades obtained by the PSS and the GSA. The greatest gap in final grades arose between the first and second place of the PSS (subjects B and F), showing a difference totaling 40%.

[9] The PSS' interpretation of subjects' final grades relates to the number of years the subject has been taking instrumental lessons. The GSA was solely intended for research purposes, it was not made clear how the researcher interpreted subjects' final grades. EEs were instructed to use the RCM's grading practices where any grade below 60% indicates a failing mark.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 22. Final Grades Obtained with Each Method of Assessment

Subject	Piano Sight-reading Scale	Gilman's Scoring Algorithm	Expert Examiners
A	18.33	71.00	63.33
B	68.33	84.60	93.66
C	11.67	15.29	55.66
D	26.67	66.31	72.66
E	26.67	64.59	76
F	28.33	71.18	94
G	10.00	49.07	69
H	25.00	66.45	82.34

Note. All grades are calculated out of 100 points.

This is especially striking since the differences between the final grades of the first and second place were of 13.42% with the GSA (subjects B and F) and 0.34% with EEs (subjects F and B). No other important differences occurred in the final grades with the PSS, the other distances ranging from 1.67% to 6.67%. The next important interval within the final grades arose between the seventh and eighth place with the GSA (subjects G and C) totaling 33.78%. Conversely, the differences between the final grades of the seventh and eighth place were of a mere 1.67% with the PSS (subjects C and G) and 7.67% with EEs (subjects A and C). No other significant divergence occurred in the grades with the GSA, these ranging from 0.07% to 15.52%. No considerable intervals arose from the evaluation by EEs with final grades ranging from 0.34% to 11.32%. Nevertheless, some consistency was observed, final grades obtained by EEs were always higher than both the PSS and the GSA, and those of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

GSA are consistently higher than the ones determined by the PSS. However, the important hiatuses in the final grades obtained with the PSS and the GSA, when no such discrepancy was found with EEs, demonstrates not only the inconsistency within the PSS and the GSA's grading but also the variations existing between the three methods.

A comparison of the final grades determined by the three methods of assessment for each subject also yielded interesting observations. The difference between the final grades determined by the PSS and the GSA for each subject provided high disparities, ranging from 37.92% for subject E to 52.67% with subject A, with two exceptions being subject B with a difference of 16.27% and subject C with a difference of 3.62%. The disparities between the final grades determined by the PSS and EEs for each subject was exceedingly high, ranging from 43.99% for subject C to 65.76% for subject F, with the exception of subject B with a difference of 25.33%. The disparities between the final grades determined by the GSA and EEs for each subject were less significant, ranging between 7.67% for subject A and 22.82% for subject F, with the exception of subject C with a difference of 40.37%. Another noteworthy observation resulted from this comparison, for every subject, the final grade determined by EEs was higher than both the PSS and the GSA, with the exception of subject A whose final grade was 7.67% higher with the GSA than with EEs. Through this comparison it was found that an important divergence exists between the final grades obtained with the PSS and both those determined by the GSA and EEs. And even though consistency was found in the subjects' final grades with each method, one noteworthy discrepancy occurred which further leads to the conclusion that these methods are dissimilar in their grading of subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conclusion

The comparison of final grades obtained from the assessment of subjects with the three methods resulted in many interesting observations. The examination of each method's final grades from the highest to the lowest determined that large gaps occurred in the assessment with the PSS and the GSA when no such variation was found with EEs. However, consistency was observed, where grades obtained by EEs were always higher than those determined by both the GSA and the PSS. Lastly, the comparison of final grades obtained by the three methods in relation to subjects, found great intervals to exist between the PSS' final grades and those obtained by the GSA and EEs. Furthermore, through this comparison of subjects' final grades, it was found that the EEs' marked higher than both the GSA and the PSS with one exception. The findings derived from the comparison of final grades further lead to the conclusion that these methods are dissimilar in their grading of subjects.

4.2.3. Subject Placement

Through the comparison of subject placement from best to weakest sight-reading performance many interesting observations were put forth (see Table 23 for subject ranking with each method of assessment, first place being the best performer and eighth place being the weakest). This comparison was divided into two sections: 1) a comparison of subject placement examining the extent to which subjects were placed in different ranks with the three methods; and 2) a comparison of subject placement in relation to music reading research.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 23. Ranking with the Three Methods of Assessment

	Piano Sight-reading Scale		Gilman's Scoring Algorithm		Expert Evaluation	
	Subject	**Grade**	**Subject**	**Grade**	**Subject**	**Grade**
First place	B	68.33	B	84.60	F	94
Second place	F	28.33	F	71.18	B	93.66
Third place	D & E	26.67	A	71.00	H	82.34
Fourth place			E	64.59	E	76
Fifth place	H	25.00	H	66.45	D	72.66
Sixth place	A	18.33	D	66.31	G	69
Seventh place	C	11.67	G	49.07	A	63.33
Eighth place	G	10.00	C	15.29	C	55.66

When comparing subject placement with each of the three methods of assessment many divergences were observed. Firstly, not one instance occurred where a subject was placed in the same rank with all three methods of assessment. The most comparable placement occurred with subject E who ranked fourth with both the GSA and the EEs, with the PSS subject E tied with subject D for third and fourth place. Then there are four instances where subjects were ranked in the same position by two of the three methods of assessment: subject B ranked first with both the PSS and GSA; subject F ranked second with both the PSS and GSA; subject H ranked fifth with both the PSS and GSA; and subject C ranked eighth with both the GSA and EEs. Interestingly, three out of the four instances where subjects ranked in the same position by two methods occurred between the PSS and the GSA, in these same occurrences, EEs ranked subjects between one to two places apart

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from the other two methods. Through this comparison of subject placement, it was found that for the most part, subjects placed two to three positions apart with the different methods. However, two instances occurred where subjects were placed with greater severance by the three methods: subject D tied for third and fourth place with the PSS, then was ranked sixth with the GSA and fifth with EEs for a difference of four positions; and subject A ranked sixth with the PSS, third with the GSA, and seventh with EEs for a difference of five positions. The comparison of subject placement demonstrated that even though the majority of subjects were placed in ranks of relatively close proximity by the three methods, no subject was placed in the same rank by all three methods, and although few large divergences occurred, these at times were as great as five out of eight positions.

The three assessment methods presented in this study are often used by researchers to categorize good and poor readers: subjects in the top 50% become the good readers and the bottom 50%, the poor readers. In our experimentation, subjects ranking first, second, third, and fourth would be considered good readers and those ranking fifth, sixth, seventh, and eighth would be considered poorer readers. Our analysis shows that many divergences would then occur among the subjects placed in each group depending on the assessment method being used (see Table 24). In view of this placement determined by the three methods, three instances occurred when subjects would have placed in different groups, good readers and poor readers, depending on the method of assessment. Subject D would have been considered a good music reader with the PSS, and a poor reader with the GSA and EEs. Subject H would have been considered a poor reader by both the PSS and GSA whereas a good reader with EEs. And subject A would have been considered a poor reader by both the PSS and EE whereas a good reader with the GSA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 24. Placement as Good or Poor Reader with Each Method of Assessment

Group	Piano Sight-reading Scale	Gilman's Scoring Algorithm	Expert Examiners
Good Readers	Subject B Subject F **Subject D** Subject E	Subject B Subject F **Subject A** Subject E	Subject F Subject B **Subject H** Subject E
Poor Readers	**Subject H** **Subject A** Subject C Subject G	**Subject H** **Subject D** Subject G Subject C	**Subject D** Subject G **Subject A** Subject C

This implies that 37.5% of the subjects participating in this research project would have varied in their ranking as good or poor readers depending on the assessment used to classify them. These observations indicate that music reading research using one of these three methods of assessment to rank subjects would place their participants in very different groups depending on the method used.

4.2.4. Conclusion

The difference in the assessment of eight subjects resulting from the grading with the PSS, the GSA, and the EEs are apparent when looking at the grade deductions, the final grades, and subject placement. The comparison of grade deductions resulting from the assessment with the three methods found that pitch errors were consistently graded most severely by the PSS than by the GSA and EEs. It was also found that the GSA and EEs graded key signature errors very differently, the GSA grading such pitch errors most leniently, whereas EEs emphasized these as very important errors. Rhythm errors did not differ as widely as pitch errors, however, interesting disparities were found. In addition, the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

GSA consistently found performances to contain rhythm errors, however, did not necessarily stress the same type of rhythm errors as EEs. Finally, articulation and expression errors were the predominant types of other errors, however, assessments often differed in the other errors found, and at times the PSS deduced important marks for such errors when these were not stressed by EEs and not considered by the GSA. The comparison of final grades also put forth many noteworthy observations. Through the examination of each method's final grades, large gaps were found among the marks obtained with the PSS and the GSA, however, no such discrepancy was found with EEs. However, consistency was found when observing the final grades from the highest to the lowest, where EEs, without fail, graded each rank higher than the GSA and the PSS. And even though some consistency was found in the comparison of final grades in relation to individual subjects, one important discrepancy occurred which leads to the conclusion that these methods are divergent in their grading of subjects. The comparison of subject placement revealed that even though the bulk of participants ranked in relatively contiguous placements with the three methods, no subject was positioned equally by all three methods, and even though relatively great variances rarely occurred, these were at times as large as five out of eight positions. And finally, the comparison of subject grouping as good or poor music reader found that 37.5% of subjects in this research project would have ranked differently depending on the method employed to assess them. All of these findings lead to the assumption that these three methods do not necessarily provide analogous assessments of sight-reading performances.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5

5. Discussion

The purpose of this study was to determine if three sight-reading assessment methods used in music reading research provide equivalent analyses when measuring the same performances. In chapters three and four, the three methods' assessment of eight piano students' performances was first presented, and then compared. The present chapter provides a discussion and conclusion for the comparison of the three methods' assessments. This chapter will first discuss the comparison of: 1) assessment procedures; and 2) subject assessment. This chapter will be followed by a conclusion for this study, providing a summary of our research and its findings and offer comments for further research in the field.

5.1. Comparison of Assessment Procedures

The three assessment methods were first compared in relation to their assessment procedures. It was found that these differed significantly in relation to: 1) the type of error considered by each method; 2) their allocation of marks; and 3) the length and complexity required for the undertaking of each of the assessments. The present section will discuss the disparities existing in the three methods' assessment procedures.

5.1.1. Types of Errors

The comparison of the types of errors considered by the PSS, the GSA and EEs presented many noteworthy disparities. It was found that the PSS and EEs have many similarities; however, these differed greatly from the GSA. The PSS and the EEs were found

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to have more detailed error classifications, taking into consideration numerous aspects, such as pitch, rhythm, timing, articulation, expression, etc, while the GSA provided a more limited classification of errors, measuring only timing inaccuracies as well as pitch errors. Interestingly, in previous research, it was found that many different types of errors are commonly made when reading musical notation such as pitch, rhythm, dynamics, articulation, and timing (Jaarsma, Ruijssenaars, & Van Den Broek, 1998; Whitlock, 2002). This would necessarily require that a broad range of errors be considered when assessing music sight-reading performance. Furthermore, the review of different definitions of music reading and sight-reading presented in the first chapter of this paper shows that these tasks require the interpretation and performance of the many possible notational elements, or musical symbols, found in music. Ely and Rashkin (2005) list a few of these as "notes, expression markings, tempo indications, etc." (p. 284) and Lehmann and McArthur (2002) reaffirm that music sight-reading requires an appropriate tempo and proper expression. Again, this would point towards the assessment of sight-reading performance requiring an examination of the many elements required of this task. In light of these observations made from music reading research literature, it can be established that the limited errors assessed by the GSA are not totally appropriate for the measurement of sight-reading. The more thorough and inclusive error classifications used by the PSS and EEs would necessarily be more appropriate in relation to their consideration of the many types of errors possible in piano sight-reading performance.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.1.2. Allocation of Marks

The allocation of marks to the different types of errors was another palpable inconsistency found when comparing the procedures used by these three assessment methods. The most obvious discrepancy lies in the comparison of the PSS and the GSA. The PSS did not make a distinction between the various types of errors, but rather deducted the same mark regardless of the types of errors and used the musical measure as grading unit. On the other hand, the GSA provided a very specific and regulated mathematical algorithm to calculate pitch and rhythm errors in relation to individual notes. This algorithm was established by Gilman (2000) to take into consideration the different types of pitch and rhythm errors, deducting more marks for errors she felt were more important. Interestingly, these two very different assessment methods were both found to be problematic in their allocation of marks: the PSS for its deduction of tremendously high marks resulting from errors such as articulation and expression; and the GSA for its subtraction of negligible marks resulting from pitch errors and extremely high mark deductions for pauses and added notes and rests. Furthermore, the PSS' use of the measure as grading unit resulted in illogical mark deductions, another limitation of this method. A look at the EEs' comments on their evaluation of sight-reading performance provided a good view of the varying importance placed on the different types of errors by a music education community. The three EEs communally stressed such errors as pitch, rhythm, and timing over other types of errors such as articulation and expression. Though evaluation by EEs was not as standardized or objective as the other two methods, resulting at times in large discrepancies among the final grades, they did provide a logical mark deduction taking into consideration the differing types of errors and the amount of importance they felt needed to be placed on each of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

various types of errors. Through the comparison of the allocation of marks, it was first determined that noticeable differences exist between the three methods and that both the PSS and the GSA, created as standardized methods for the assessment of sight-reading, do not follow grading procedures that conform to the importance placed on specific errors by a prominent educational community such as the Royal Conservatory of Music. This would lead to questioning their validity not only in a modern educational setting but also for use in music reading research.

5.1.3. Length and Complexity

The length and complexity of the grading procedures necessary for each of the three methods further distinguishes them. The PSS and EEs required relatively straightforward and rapid assessment procedures, whereas the GSA called for a lengthy and complex process. In order to make a grading procedure alluring to educational and research communities, it would not only have to be valid and standardized, but its time consumption and complexity should preferably be at a minimum. Therefore, when only taking into consideration the length and difficulty of each method, in both research and education settings, the PSS and EEs' assessment procedures would be more appealing than the GSA. This could possibly explain why none of the sight-reading error quantifications discussed in this study's review of literature have been used in subsequent research projects, with the exception of Gilman's (2000) Scoring Algorithm, which was used in a later study by the same author. Furthermore, the rapidity and straightforward nature of evaluation by EEs is most likely why this is the preferred method in music education settings and why this procedure can easily be found in music reading research. This can also be said of the WFPS, which we adapted into the PSS,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which is also commonly used in both education and research settings. Therefore, the comparison of the length and complexity of undertaking these assessments determined that the GSA, because of its time consumption and intricate procedure, might not be as appealing as the PSS and EEs.

The comparison of the grading procedures used by the three methods being examined in this research project demonstrated not only that these differ greatly in the types of errors being considered, the allocation of marks to the varying types of errors, as well as the length and complexity of their assessment processes, but also leads to questioning the validity of the PSS and the GSA as these assign unwarranted mark deductions which signal a departure from the norms employed by a leading community in music education. Furthermore, the complexity and time consumption required of the GSA would most likely make it unpopular in both educational and research settings. Through this comparison, the following can be said of sight-reading performance assessment in pianists: 1) it is a complex skill and, therefore, requires the assessment of many elements such as pitch, rhythm, timing, articulation, expression, etc.; 2) the allocation of marks to individual types of errors should take into consideration the varying importance placed on such errors by experts knowledgeable in music education practices and grade accordingly; and 3) the degree of difficulty and time necessary for undertaking the assessment should be kept at a minimum.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.2. Comparison of Subject Assessment

The three methods were then compared in relation to their assessment of eight subjects' sight-reading performance on five exercises. This section will discuss the comparison of subject assessment in three parts: 1) grade deductions; 2) final grades; and 3) subject placement.

5.2.1. Grade Deductions

The comparison of grade deductions found many inconsistencies to exist between the three methods. Three important observations are here discussed: 1) grade deductions for pitch errors; 2) grade deductions for rhythm errors; and 3) grade deductions for other errors.

Pitch Errors

Pitch errors were among the predominant types of errors found in the assessment of subjects' performances; two noteworthy observations are here discussed. Firstly, in many performances where the exact same pitch errors were located, incredible disparities were found in the deductions made to the final grades. For example, variances in subtractions to the final grades resulting from pitch errors at times surpassed 60% between the PSS and the GSA. However, this can easily be explained by an examination of these two methods' allocation of marks. The PSS grades per measure, therefore, if pitch errors are found within numerous measures, grade deductions are significant, while the GSA grades per note, and furthermore, quantifies the distance between the wrong pitch and the intended pitch, resulting in minimal grade deductions. Interestingly, grade deductions by EEs were often found to be between those of the PSS and the GSA. This not only displays the disparities between these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

three methods, but also the existing limitations where grade deductions to one of the most common types of errors differ greatly.

Secondly, two important inconsistencies were observed in relation to grade deductions resulting from pitch errors by the GSA and EEs. With the GSA, subjects received negligible grade reductions resulting from pitch errors. Gilman (2000), in her explanation to her Scoring Algorithm, stated:

> [...] the graded point system accounts for situations where the pianist simply forgets to perform a sharp or flat, then the error made will be relatively small (e.g. a semitone or a tone) and will not necessarily result from poor sight-reading ability (p. 46).

This statement clearly displays that, according to Gilman, when a subject neglects to perform in the appropriate key signature, the grade deductions should be minimal. In contrast, EEs stressed the subjects' observation or neglect of the key signature. In the EEs' grading strategies and reasoning, key signature errors seemed to have more weight then any other type of pitch error. Secondly, Gilman's statement implies that performances should be judged more severely when the distance between the performed pitch and the intended pitch is greater. In our quantification of performances with the GSA, the highest grade deductions resulting from pitch errors arose from performances at the wrong octave where the subject was subtracted one point for each pitch, ensuing in 25% or more being deduced from the final grade. Interestingly, this type of pitch error did not seem as important to EEs, and was not particularly stressed in their comments. Therefore, the EEs comments on pitch errors lead us to conclude that Gilman's strategy might be problematic, or at the very least, does not conform to educational practices: where failing to observe the key signature indicates a poor

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sight-reading performance and should be graded more severely, while performance at the wrong octave should not be penalized too strictly.

Rhythm Errors

Rhythm errors were among the leading types of errors found with the three methods of assessment; two important observations are here discussed. Firstly, the GSA consistently found performances to contain rhythm errors, even when both the PSS and EEs determined performances to be faultless in relation to rhythm. Nevertheless, the location of rhythm errors among all performances with the use of the GSA is understandable considering the use of CuBase Musical Scores (CMS) for the quantification process. With the use of the CMS all rhythmic inconsistencies, no matter how minimal, were observable, and therefore, gradable. A trained musical ear might not be as systematic in its grading. However, another explanation for this inconsistency of grade deductions between the GSA and both the PSS and EEs needs to be considered: this explanation lies in a good musician's phrasing of music where, at the end of a musical phrase, slowing of the tempo at times occurs. The GSA does not account for such a normal and desirable behavior while the trained musicians, teachers, and examiners assessing performances for the PSS and EEs portion of the experiment would most likely not consider such behavior to be a rhythm error, but a natural response to musical phrasing. This second explanation calls to question the use of the GSA, a strict and possibly unmusical method of quantifying sight-reading performance.

Secondly, even though in the previous paragraph the GSA was found to be overly strict in some cases, in other instances it graded less strictly than EEs in relation to rhythm errors. Gilman (2000) made the following statement to explain her Scoring Algorithm:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> [...] a pianist who performs a note fractionally too long will score higher than a pianist who performs the note for double its annotated duration [...] a pianist who makes 3 attempts at performing a note should score less than a pianist who performs a note correctly on one attempt. Similarly, a pianist who pauses (rests) before performing a note correctly should score less than a pianist who performs a note correctly without pausing (pp. 46-47).

With this statement Gilman indicates that a performer who makes larger rhythm errors should be penalized more severely than those who make smaller rhythm errors, and a performer who pauses and adds notes should also be graded more rigorously. In our quantification with the GSA, subjects lost the most marks for pauses, hesitations, added notes and rests, and false starts. However, in the few instances where subjects performed without taking into consideration the rhythm, performing all notes at the same length whether they were half-notes, quarter-notes, or eighth-notes, performances were not necessarily graded very severely if the subject did not pause, hesitate, or add notes. This was in contrast to the evaluation by EEs, where this type of rhythmic error was stressed and such performances graded more strictly. This leads us to conclude that even though Gilman's statement rings true in concept, her algorithm does not necessarily provide accurate grading of piano sight-reading performances in all instances when comparing it to EEs evaluations of the same performances.

Other Errors

The most predominant types of other errors found within performances were articulation and expression: one important observation is here discussed. As the GSA did not

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

take into consideration errors other than pitch and rhythm, no other errors were determined by this method. However, both the PSS and EEs did consider such errors as articulation and expression, nevertheless, in numerous instances, such errors were found solely by one of the methods and not by the other. Furthermore, great disparities occurred in the grade deductions to other errors by the PSS and EEs. EEs tended to place little importance to errors such as articulation and expression and it can be assumed that few grades were deducted because of them. Conversely, the PSS did not distinguish between the various types of errors, resulting in significant mark deductions for articulation and expression errors. Taking into consideration that the GSA did not include such errors in its grading procedure and that EEs did not seem to place much importance to these errors leads to question the PSS in relation to its significant grade deductions to other errors.

Conclusion

The comparison of grade deductions in relation to pitch errors, rhythm errors and other errors found many conspicuous discrepancies to exist between the three methods. Grade deductions to pitch errors were often found to differ immensely, the PSS deducing significant marks, the GSA negligible marks, whereas the EEs graded somewhere in the middle. A comparison of the GSA and EEs' grading of pitches in relation to key signature errors and performance at the wrong octave, leads to question the GSA's grading procedure as it does not conform to the practices of a leading educational community. In the comparison of grade deductions resulting from rhythm errors, it was found that the GSA always found such errors within performances, even when no such errors were found by the PSS and EEs. Furthermore, in other instances, the GSA was found to grade less strictly than the EEs, again leading to question the GSA for its non-musical grading procedure and its

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

disconnect to educational practices. Finally, the comparison of other errors found that even though articulation and expression errors were commonly found within the PSS and EEs grading, these varied significantly not only in relation to when they were found, but also in relation to grade deductions, leading to question the PSS' significant grading of other errors. Through this comparison of grade deductions, both the PSS and GSA were found to stray from conventional practices used by experts knowledgeable in music education leading to question their validity within a music education context.

5.2.2. Final Grades

The three methods of assessment were then compared in relation to the final grades obtained for each rank as well as for each subject. This section will discuss the following: 1) observations of the final grades from the highest to the lowest; and 3) observations between the three methods' final grades in relation to each subject.

Even though a comparison of final grades is to some extent limited, as these methods' marking of performances were not intended for use in the same context, their comparison was still able to provide interesting findings. Firstly, when observing final grades from the highest to the lowest, the distance between the eight ranks' final grades was relatively small, nevertheless, on two occasions significant distances occurred between the final grades: the first between the first and second place of the PSS; and the second, occurred between the seventh and eighth place of the GSA. Interestingly, no such difference occurred in the EEs' final grades. The fact that important differences within the PSS and the GSA's final grades did not occur between the same ranks, and that no such gap was found in the EEs' final grades, not only demonstrates these methods' discrepancies, but also leads to question both

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the PSS and GSA as they obviously do not follow grading practices used by this music education community. Nevertheless, some consistency was found where EEs always graded subjects higher than both the PSS and the GSA and that the GSA's final grades were constantly higher than the PSS. This finding, interestingly, demonstrates that some consistency exists between these three methods. Finally, final grades were compared between the three methods in relation to individual subjects, and again grades were found higher with the PSS, second highest with the GSA, and lowest with the PSS. However, one exception occurred where the GSA graded one subject less strictly than EEs. Even though, again this finding demonstrates consistency, this one exception points towards discrepancies between the three methods. It can, therefore, be concluded that although some consistency was found in final grades obtained by the three methods, again irregularities and significant gaps were observed signifying the disparities between the methods.

5.2.3. Subject Placement

The order in which subjects ranked with each of the three methods of assessment was then compared. This section will examine: 1) the degree to which subjects were ranked in differing positions by the three methods; and 2) subject placement in relation to music reading research.

Through the comparison of subject placement with each of the three methods of assessment, two noteworthy observations were made. Firstly, even though it was found that not one subject was placed in the same position by all three methods, in one instance, one subject was ranked equivalently by two methods, and tied with another subject for the same rank with the third method. Furthermore, in four instances, the same subject was ranked

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

equivalently by two methods and was placed one to two ranks apart by the third method. Interestingly, three of these four instances occurred between the PSS and the GSA. This would lead to the conclusion that even though these methods have some disparities, they do seem to rank subjects in relatively close proximity, nevertheless, it does appear as though the PSS and the GSA have more similarities than EEs. On the other hand, two subjects ranked four and five places apart by the three methods. These significant distances between the placements of two out of this study's eight subjects definitely points towards important differences existing between the assessments obtained by these methods. It can therefore, be concluded that even though the majority of this study's subjects ranked in close proximity by all three methods, one-quarter of the participants were placed in very different ranks leading to the question of equality in the assessments determined by these methods.

In music reading research these three methods are commonly used to rank participants as good or poor music readers. In the previous chapter, we showed how in music reading research, participants ranking in the top half would be considered good music readers, while those ranking in the bottom half would be considered poor readers. When this situation was used to examine the placement disparity of the eight subjects in our research project, it was found that more than one-third of our participants would have differed in their ranking as good or poor readers depending on the assessment used to classify them. Our findings, therefore, would lead to the conclusion that subjects would be placed in different groups in relation to their sight-reading ability depending on the assessment method employed by researchers. This conclusion is of great importance not only to music researchers but also to music education specialists, as it demonstrates the blatant incongruent

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

results attained by three commonly used methods and places doubt on their assessment for piano sight-reading performance.

Through the comparison of subject placement by the three methods of assessment it was found that even though, the majority of subjects ranked in relatively close proximity, one-quarter of this study's participants differed significantly in their placement. In addition, more than one-third of this study's participants would have differed in relation to their classification as a good or poor music reader depending on the method used to assess them. Therefore, even though for the most part, subject placement was somewhat equivalent, the disparities found in subject ranking lead to the conclusion that these methods definitely do not provide analogous assessments of sight-reading performance and should not be used interchangeably in music reading research.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6. Conclusion

A large number of music reading research projects examine the complex task of music reading by comparing the differences between good and poor music readers, whether the research are in eye movements, cognitive strategies, or teaching approaches. After reviewing numerous studies, it was evident that many different methods were being used to rank subjects as either good or poor readers. We, therefore, felt it was necessary to first determine if the various methods used by music research and education communities to assess sight-reading performance provided analogous rankings of subjects in order to pursue further study in the field. Three commonly used methods, the Watkins-Farnum Performance Scale, Gilman's Scoring Algorithm, and Expert Examiners, were selected for comparison. These three methods were to be used for the assessment of new piano sight-reading performances, therefore, the WFPS, developed for use with wind instrumentalists, was adapted into the Piano Sight-reading Scale by composing new exercises, and then adjusting the grading system and error characterization. These three methods were then compared in two parts: 1) their assessment procedures; and 2) their assessment of eight piano students' sight-reading performances of five exercises. It is important to note that in this research project the terms measurement, quantification, and evaluation were defined to be used in very specific contexts, however, these terms are often used interchangeably and one needs to be careful as it is difficult to precise the exact differences between them.

This study found that the assessment methods differed in both of the comparisons made. Firstly, their assessment procedures differed in relation to the types of errors they considered, their allocation of marks to each of the various types of errors, as well as their time consumption and complexity. Through this comparison, it was found that the PSS and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the GSA's mark deduction procedures did not follow the norms employed by the three EEs, respected examiners with the Royal Conservatory of Music, leading to questioning their validity.

The comparison of the performance assessment found that these methods differed in grade deductions, final grades, and subject placement. In the comparison of grade deductions to pitch errors, at times very different marks were deducted for identical errors. Furthermore, important discrepancies were observed between the GSA and EEs, in the assessment of pitch errors, the GSA deducted negligible marks resulting from key signature errors when these were severely stressed by EEs, and the GSA deduced the highest pitch marks for performances at the wrong octave when these types of pitch errors did not seem to be as significant for EEs. In the assessment of rhythm errors, it was found that the GSA always found performances to contain such errors, even when these were not found by the other two methods. However, at times, subjects received little grade deductions with the GSA for lacking to perform the correct rhythms if their performance was steady, and without hesitations or added notes. Conversely, EEs stressed the lack of correct rhythms and graded these performances severely.

The comparison of final grades found many divergences among the three methods. The contrast of final grades from the highest to the lowest found noteworthy gaps within the final grades obtained by the PSS and the GSA which did not exist with EEs. The comparison of final grades between the three methods found some consistency in the grading of individual subjects: EEs' final grades always being highest, the GSA second highest, and the PSS obtained the lowest grades. However, one exception occurred where the GSA graded one subject higher than EEs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The comparison of subject placement yielded the most important and noteworthy findings for this research project. It was found that these three methods of sight-reading assessments did not place the eight subjects of this research project in equivalent ranks. In a few instances the same subject was placed in equivalent ranks by two methods, and for the most part, subjects ranked in positions relatively close together by the three methods. However, not only did none of the subjects place in the same rank by all three methods, more than one third of subjects would have been placed in different groups, good or poor, in relation to their sight-reading performance depending on the method used to assess them.

This thesis clearly demonstrated that these three sight-reading assessment methods used by researchers to rank subjects are not comparable and should therefore not be used with the same intent. Through this comparison of these three methods the following can be said of sight-reading performance assessment at the piano. Firstly, sight-reading is a complex process necessarily requiring a well developed and thoroughly researched method for its assessment. Both the PSS and the GSA were found to have grading procedures which do not conform to the grading practices of a major examination institution, in relation to the types of errors being assessed and the allocation of marks in relation to those errors. A proper method for the assessment of this skill would take into consideration the multi-faceted nature of sight-reading, encompassing the many elements necessary, such as pitch, rhythm, timing, articulation, expression, etc. It would grade each type of errors according to its importance within the task in question. Therefore, when developing such as method, many experts knowledgeable in the assessment of piano sight-reading should be consulted in order to develop an appropriate grading procedure that reflects educational practices. Finally, it should be a method accessible to both researchers and the education community by its

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

straightforward and rapid procedure. This study clearly demonstrated not only the need for further research in the field of music reading but also the immediate requirement for a standardized and valid method for the assessment of piano sight-reading performance.

Nevertheless, it is important to keep in mind this study's limitations. Firstly, this study used a small number of subjects to be assessed by the three methods under examination. Secondly, both the WFPS and the GSA made use of a metronome to introduce a tempo to subjects and we did not as we felt it would disrupt our young students who might not have had experience sight-reading with a pre-set tempo. In addition, we feel that an integral part of sight-reading ability is the selection of an adequate tempo, having students perform with a metronome would make it impossible to judge this essential requirement of a good sight-reading performance. Furthermore, some of the studies using the WFPS to measure sight-reading performances had evaluators practice with the sight-reading error classification before undertaking the actual experiment. Because of timing issues, our testers were asked to carefully review the instructions, and any concerns or questions were answered, but they did not have the opportunity to practice the assessment process. And finally, the EEs chosen for this experiment only represented one music education community. These limitations should be carefully considered.

The next and most important step for future research in music reading is the development of a standardized test for the assessment of piano sight-reading. This study is a starting point for further research on the assessment of music sight-reading in pianists providing many noteworthy findings. It determined that sight-reading requires the processing of many elements, such as pitch, rhythm, expression, articulation, etc., and that all these elements should be included in a proper assessment of this skill. The allocation of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

marks to the various types of errors should consider the differing degrees of importance placed on individual errors by experts knowledgeable in music education practices. Finally, the complexity and time necessary for the assessment procedure should be minimal. The many findings derived from the comparison of these three methods provide a foundation for the development of a formal and standardized piano sight-reading test. This new method would not only be a great addition to music reading research, providing a standardized and validated assessment tool, but could also serve the music education community which, as of yet, has no instrument for the measurement of sight-reading in young piano students. This research project broadened our understanding of three methods implemented for the assessment of sight-reading performance in music reading research, while providing further knowledge on this complex skill and paving the way for the construction of necessary methods for the assessment of sight-reading in pianists.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

References

Abeles, H. F. (1973). A facet-factorial approach to the construction of rating scales to measure complex behaviors. *Journal of Educational Measurement, 10*(2), 145-51.

Agay, D. (2004). *The art of teaching piano: The classic guide and reference book for all piano teachers.* New York: Yorktown Music Press.

Anderson, J. N. (1981). Effects of tape-recorded aural models on sight-reading and performance skills. *Journal of Research in Music Education, 20*(1), 23-30.

Babin, A. (2005). *Music conservatories in Canada and the piano examination system for the preparatory student: A historical survey and comparative analysis.* (Master's thesis, University of Ottawa).

Baker-Jordan, M. (2004). *Practical piano pedagogy: The definitive text for piano teachers and pedagogy students.* Florida: Warner Bros. Publications.

Bamberger, J. (1999). Learning from the children we teach. *Bulletin of the Council for Research in Music Education, 142,* 48-74.

Bastien, J. W. (1995). *How to teach piano successfully* (third edition). San Diego: Neil A. Kjos Music Company.

Bennett, E., & Capp, H. (1969). *Complete series of sight-reading and ear tests.* Mississauga, ON: Frederick Harris Music.

Berlin, B. (1984). *The A.B.C. of piano playing: An easy method for beginners.* Mississauga, ON: Frederick Harris Music Co.

Berlin, B. (1986). *Four star sight reading and ear tests: Daily exercises for piano students* (Rev. ed.). Oakville, ON: Frederick Harris Music.

Berlin, B. (1984). *The ABC of piano playing: An easy method for beginners* (Rev. ed.). Mississauga, ON: The Frederick Harris Music Co.

Bloch, H., Dépret, A., Gallo, A., Garnier, Ph., Gineste, M.-D., Leconte, P., Le Ny, J.-F., Postel, J., Reuchlin, M., & Casalis, D. (2002). *Dictionnaire fundamental de la psychologie* (2 vols.). Paris: Larousse.

Boyle, J. D. (1968). The effects of prescribed rhythmical movements on the ability to sight-read music. (University of Kansas, 1969). *Dissertation Abstract International*, 29(07), AAT 6817359.

Boyle, J. D. (1970). The effect of a program for teaching sight-reading in junior high school training bands. *Journal of Band Research, 7*(1), 7-15.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Boyle, J. D. (1992). Evaluation of music ability. In R. Colwell (Ed.), *Handbook of research in music reading and learning* (pp. 247-265). New York: Schirmer Books.

Bowles, R. W. (1971). *Belwin-Mills singing achievement test*. Melville, NY: Belwin-Mills.

Bradley, D., & Tobin, J. R. (1947). *Sight-reading made easy: A complete graded piano course*. London: Strainer & Bell.

Burrows, T. (1999). *How to read music: Music reading made simple*. New York: St. Martins Press.

Camp, M. W. (1992). *Teaching piano: The synthesis of mind, ear and body*. Los Angeles: Alfred Publishing Co.

Colman, A. M. (2006). *A dictionary of psychology: Oxford reference online*. Oxford: Oxford University Press. University of Ottawa. http://www.oxfordreference.com.proxy.bib.uottawa.ca/views/ENTRY.html?subview=Main&entry=t87.e276

Colwell, R. J., & Richardson, C. (Eds.) (2002). *The new handbook of research on music teaching and learning: A project of the Music Educators National Conference.* New York: Oxford University Press.

Colwell, R. J. (1991). Evaluation. In R. J. Colwell (Ed.), *Basic concepts in music education II* (pp. 247-278). Niwot, CO: University Press of Colorado.

Christ, W. B. (1953). The reading of rhythm notation approached experimentally according to techniques and principles of word reading. (Indiana University, 1953). *Dissertation Abstract International, 14*(4), AAT 0007538.

Clark, F., & Goss, L. (1993). *The music tree: A plan for musical growth at the piano*. Miama, FL: Summy-Birchard.

Craige, M. A. (1993). A survey of the instruction of sight-reading skills to undergraduate piano majors in selected NASM colleges and universities. (University of North Texas, 1993). *Dissertation Abstract International, 54*(05), AAT 9326622.

Deutsch, D. (Ed.). (1998). *The psychology of music* (2nd ed.). San Diego: Academic Press.

Dodson, T. A. (1983). Developing music reading skills: Research implications. *Update: Applications of Research in Music Education, 1*(4), 3-6.

Eaton, J. L. (1978). *A correlation study of keyboard sight-reading facility with previous training, note-reading, psychomotor, and memorization skills*. (Doctoral dissertation, Indiana University).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Elkan, I. (1948). *Piano sight-reading.* Boston: Carl Fischer.

Elliott, C. A. (1982a). The identification and classification of instrumental performance sight-reading errors. *Journal of Band Research, 18*(1), 36-42.

Elliott, C. A. (1982b). The relationships among instrumental sight-reading ability and seven selected predictor variables. *Journal of Research in Music Education, 30*(1), 5-14.

Ely, M. C., & Rashkin, A. E. (2005). *Dictionary of music education: A handbook of terminology.* Chicago: GIA Publications, Inc.

Faber, N, Faber, R., & McArthur, V. (2001). *Piano adventures.* Fort Lauderdale, FL: FJH Music Company Inc.

Farnum, S. E. (1969). *The Farnum string scale: A performance scale for all string instruments.* Winona, MN: Hal Leonard Music.

Furneaux, S., & Land, M. F. (1997). The role of eye movements during music reading. In A. Gabrielsson (Ed.), *Proceedings of the 3rd Triennial ESCOM Conference* (pp. 210-214). Uppsala, Sweden: Uppsala University.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings: Biological Sciences, 266*(1436), 2435-2440.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music, 31*(3), 221-272.

Galyen, S. D. (2005). Sight-reading ability in wind and percussion students: A review of recent literature. *UPDATE Application of Research in Music Education, 24*(1), 57-70.

Gaynor, J. (1995). *Music reading comprehension: The effect of Aid to Chunking and Melodic Predicting on sight reading performance achievement of secondary school instrumental music students.* (Doctoral dissertation, University of San Francisco).

George, W. E. (1980). Measurement and evaluation of musical behavior. In D. A. Hodges (Ed.), *Handbook of music psychology* (pp. 291-340). Lawrence: National Association for Music Therapy.

Gilman, E. R. (2000). *Towards an eye-movement model of music sight-reading.* (Doctoral dissertation, Nottingham University).

Gilman, E., Underwood, G., & Morehen, J. (2002). Recognition of visually presented musical intervals. *Psychology of Music, 30,* 48-57.

Goolsby, T. (1989). Computer application to eye movement research in music reading. *Psychomusicology, 8* (2), 111-126.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

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

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

Gordon, S. (2001). *Etudes for piano teachers: Reflections on the teacher's art.* New York: Oxford University Press.

Gromko, J. E. (2004). Predictors of music sight-reading ability in high school wind players. *Journal of Research in Music Education, 52*(1), 6-15.

Gudmundsdottir, H. R. (2003). Music reading errors of young piano students. (McGill University, Montreal, 2003). *Dissertation Abstract International, 65*(02), AAT NQ88481.

Guhl, L. (1989). *The Magic Reader.* San Diego: Neil A. Kjos Music Company.

Gunter, T., Schmidt, B., & Besson, M. (2003). Let's face the music: A behavioral and electrophysiological exploration of score reading. *Psychophysiology, 40,* 742-751.

Hahn, L. B. (1985). Correlations between reading music and reading language, with implications for music instruction (notation). (University of Arizona, 1985). *Dissertation Abstract International, 46*(09), AAT 8525597.

Haley, K. A. (1999*). Watkins-Farnum revisited: Application of the Rasch model to measures of musical performance.* (Doctoral dissertation, Boston College).

Hodges, D. A. (1980). *The handbook of music psychology.* Lawrence, KS: National Association for Music Therapy.

Hodges, D. A. (1992). The acquisition of music reading skills. In R. Colwell (Ed.), *Handbook of research in music reading and learning* (pp. 466-471). New York: Schirmer Books.

Hodges, D. A. (Ed.) (1996). *The handbook of music psychology* (2nd ed.). St. Louis, MO: MMB Music.

Jaarsma, B. S., Ruijssenaars, A., & Van Den Broek, W. (1998). Dyslexia and learning musical notation: A pilot study. *Annals of Dyslexia, 48,* 137-154.

Jacobson, J. M. (2006). *Professional piano teaching: A comprehensive piano pedagogy textbook for teaching elementary-level students* (Edited by E. L. Lancanster). Los Angeles: Alfred Publishing Co.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Johnson, T. A. (2001). *Right@Sight* (Rev. by Caroline Evans). London: Edition Peters.

Kember, J. (2004). *Sight-reading*. Mainz, Germany: Schott Musik.

Killian, J. N., & Henry (2005). A comparison of successful and unsuccessful strategies in individual sight-reading preparation and performance. *Journal of Research in Music Education, 53*(1), 51-65.

Kinsler, V., & Carpenter, R. H. S. (1995). Saccadic eye movements while reading music. *Vision Research 35*(10), 1447-1458.

Kopiez, R., Lee, J. I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research, 8*(1), 97-120.

Lafferty, L. (2002). *The visual effect of embouchure, appearance, and gender on solo clarinet performance adjudication.* (Doctoral dissertation, Kent State University).

Lehmann, A. C., & Ericsson, K. A. (1993). Sight-Reading ability of expert pianists in the context of piano accompanying. *Psychomusicology, 12,* 182-195.

Lehmann, A., & McArthur, V. (2002). Sight-Reading. In R. Parncutt, & G. E. McPherson (Eds.), *The Science and Psychology of Music Performance* (pp. 135- 150). New York: Oxford University Press.

Levy, K. L. M. (2001). Music readers and notation: Investigation of an interactive model of rhythm reading. (University of Iowa, 2001). *Dissertation Abstract International, 62*(12), AAT 3034122.

MacKnight, C. B. (1975). Music reading ability of beginning wind instrumentalists after melodic instruction. *Journal of research in music education, 23*(1), 23 -34.

McKenzie, C., Vaneerd, D. L., Graham, E. D., & Huron, D. B. (1986). The effect of tonal structure on rhythm in piano performance. *Music Perception, 4*(2), 215-225.

McPherson, G. E. (1994). Factors and abilities influencing sightreading skills in music. *Journal of Research in Music Education, 42*(3), 217-231.

Pace, R. (1954). *The Robert Pace piano series.* King of Prussia, PA: Theodore Presser Co.

Petzold, R. G. (1960). The perception of music symbols in music reading by normal children and by children gifted musically. *Journal of Experimental Education, 28,* 271-319.

Rabin, R. (1995). *At the beginning: Teaching piano to the very young child.* New York: Schirmer Books.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rogers, G. L. (1996). Effect of colored rhythmic notation on music-reading skills of elementary students. *Journal of Research in Music Education, 44*(1), 15-25.

Royal Conservatory of Music (2001). *Royal Conservatory of Music piano syllabus.* Mississauga, ON: Frederick Harris Music.

Rubinstein, B. (1950). *The pianist's approach to sight-reading and memorizing.* Boston: Carl Fischer.

Salis, D. L. (1977). *The identification and assessment of cognitive variables associated with reading of advanced music at the piano.* (Doctoral dissertation, University of Pittsburgh).

Schmoll, A. (1912). *Nouvelle méthode de piano théorique, pratique et récréative.* Paris: Édition A. Schmoll.

Scripp, L. R. (1995). The development of skill in reading music. (Harvard University, 1995). *Dissertation Abstract International, 56* (06), AAT 9534633.

Serafine, M. L. (1988). *Music as Cognition.* New York: Columbia University Press.

Shanet, H. (1971). *Learn to Read Music.* New York: Simon & Schuster

Sloboda, J. (1978). The psychology of music reading. *Psychology of Music, 6,* 3-20.

Sloboda, J. (1988). *L'esprit musicien: La psychologie cognitive de la musique.* Bruxelles : Mardaga.

Sloboda, J. (2005). *Exploring the musical mind: Cognition, emotion, ability, function.* New York: Oxford University Press.

Snyder, S., & Sheehan, R. (1992). The Rasch measurement model: An introduction. *Journal of Early Intervention, 6*(1), 87-95.

Starr, C. (2000). *The music road: A journey in music reading.* Miami, FL: Summy-Birchard.

Stauffer, D. W. (2005). Learning to read music fluently. *Music Educators Journal, 92*(1), 21-22.

Steelman, L. (1998). *Music reading for keyboard: The complete method.* Musicians Institute Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Stewart, D. (1999). *The musician's guide to reading and writing music* (2nd ed.). Milwaukee: Backbeat Books.

Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage, 20*(1), 71-83.

Thompson, J. (2005). *John Thompson's modern course for the piano* (Rev. ed.). Willis Music.

Thompson, S., & Lehmann, A. C. (2004). Strategies for sight-reading and improvising music. In A. Williamon (Ed.), *Musical Excellence* (pp. 143-159). New York: Oxford University Press.

Thompson, W. B. (1985). Sources of individual differences in music sight-reading skills. (University of Missouri, 1985). *Dissertation Abstract International, 47*(02), AAT 8607243.

Thompson, W. F., Diamond, C. T. P., & Balkwill, L. (1998). The adjudication of six performances of a Chopin etude: A study of expert knowledge. *Psychology of Music and Music Education, 26,* 154-174.

Udtaisuk, D. B. (2005). A theoretical model of piano sightplaying components. (University of Missouri-Columbia, 2005). *Dissertation Abstract International, 67*(01), AAT 3204275.

Uszler, M., Gordon, S., & Mach, E. (1991). *The well-tempered keyboard teacher.* New York: Schirmer Books.

Watkins, J. (1942). *Objective measurement of instrumental performance.* (Doctoral dissertation, Columbia University).

Watkins, J., & Farnum, S. (1954). *The Watkins-Farnum performance scale: Form A.* Milwaukee, WI: Hal Leonard Publishing Corp.

Whitlock, M. N. (2002). The application of music learning theory concepts to a notation–based piano method. *Dissertation Abstracts International,* UMI No. 8525597.

Wiltshire, E. S. (2006). The effects of visual and aural congruence on the sight-reading of music notation. (University of Washington, 2006). *Dissertation Abstract International, 67*(07), AAT 3224312.

Wolf, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research, 5*(2), 143-171.

Wöllner, C., Halfpenny, E., Ho, S., and Kurosawa, K. (2003). The effects of distracted inner hearing on sight-reading. *Psychology of Music, 31*(4), 377-389.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendixes

Appendix A

Watkins-Farnum Performance Scale: Types of Errors

1. Pitch Errors (P)	a. A tone added or a tone omitted constitutes an error.	
	b. A tone played on the wrong pitch.	(1) Fuzzy attacks or minor irregularities in pitch during the course of an extended tone are not to be counted as errors, provided most of the note has been played on the right pitch.
		(2) If the student strikes the wrong pitch when attacking a note but correctly fingers it and immediately adjusts the lip to the correct pitch without retonguing the note, no error is to be counted. If he fingers it wrong, for example playing F sharp and then changing immediately to F natural, an error is counted.
2. Time Errors (R) for rhythm	a. Any note not given its correct value is marked wrong.	(1) A sustained note must be held within one count of the correct beat. Thus, a whole note held for three full counts is marked wrong. If held for three counts and a little more it is considered right. It must extend over into the beginning of the fourth count. If it extends past the end of the fourth count, into the beginning of the fifth it again becomes wrong. Count to yourself and mark an error if the tone stops before you start to say the word "four" or after you have started to say the word "five."
		(2) Rule (1) above is to be applied to sustained tones, half notes, dotted halves, dotted quarters in six eight, etc. Be sure that the difference between the time any sustained tone is held and the correct time for it is less than a full beat.
3. Change of Time Errors (T)	a. If there is a marked increase in tempo or decrease in tempo, all measures played in the incorrect tempo are wrong.	(1) If the increase or decrease in tempo within an exercise is less than twelve (12) beats per minute, no errors are to be marked. Before giving the tests, practice with the metronome to determine the limits and then use your judgment when administering the test.
		(2) Increase in tempo: (a) Where the increase is in excess of twelve beats per minute, mark wrong only the measure in which the increase took place. If the increase has been gradual, mark the measure wrong in which you think it passed the twelve beats per minute mark of increase. (A return to normal tempo at any time is not counted as an error.)
		(3) Decrease in tempo: (a) Where the decrease is in excess is in excess of twelve beats per minute, mark wrong all measures played at that retarded speed. (b) If the drop in speed is continued, mark four (4) measures wrong, and then, if the student seems wedded to the slower tempo, stop him and inform him, that he is going too slow. Indicate the correct tempo with the metronome and let him start again at that point. If he again drops below the limit in tempo, say nothing but score all measures wrong which are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

		played too slowly. (Follow the above procedure if the student fails to play exercise No. 9 in *alla breve* time.
4.Expression Errors (E)	a. Failure to observe any expression mark constitutes an error.	(1) The fact of the response, not the degree of the response, determines whether or not an error has been made. (a) Thus an increase in volume made when f follows p or mf indicates that the student knows the meaning of Forte and has read the symbol correctly. The fact that he has increased volume too little or too much in the judgment of the scorer shall not be counted as an error. Be convinced only that the student has seen the mark, knows what it means, and indicates so by responding to it.
		(2) Failure to observe a crescendo, decrescendo, accelerando, or ritardo constitutes a single error which is ascribed to the measure in which the sign originates.
5. Slur Errors (S)	a. A slur omitted, a tongued note slurred, a slur carried onto notes which should be tongued, or a broken slur are all counted as errors.	(1) Occasionally one finds a student who has been taught to play with an especially legato tongue. Do not mark slur errors when you first discover this, but inform him it is his responsibility to distinguish between slurred and tongued notes. After that, grade him as strictly as others. If in doubt, mark as error. The burden of proof lies with the student. See that he understands that and then score rigidly.
6. Rests (R)	a. Ignoring a rest or failure to give a rest its correct value is an error. Apply the same standards as to a sustained note (see rule 2).	
7. Holds and Pauses (R)	Holds written thus (fermata) should be treated in the same way as other expression marks (See section 4a, rule (1))	
	b. Pauses between measures, no matter how long, are not to be counted as errors; however, do not inform the student of this fact.	
	c. Pauses between notes within the measure are to be counted as errors.	
8. Repeats	a. Record an error if the pupil fails to make the repeat in No. 4 but score only the first rendition. In No. 7, score only the first rendition plus the second ending. Allow the student to complete the repeat but do not change the scoring or mark any new errors until the second ending	(1) If he should stop and ask you immediately whether he should repeat, answer, "Of course, play it exactly as written" and do not record an error. Failure to play the second ending on the repeat in No. 7 constitutes an error on the first measure of the second ending. In this case, tell the student to start on the second ending and finish.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix B

Adjudicator Question Document

1. Please read the question carefully.

2. Please highlight the answer which best represents your judgment of piano sight-reading evaluation.

3. Please provide reasoning for your answer.

Question:

If you are grading piano students on sight-reading ability and students are given exercises of four to eighteen measures, each which gradually increase in difficulty, **how many wrong measures** would you permit and still consider that the student should receive some credit?

Answer:

f. all measures

g. almost all measures

h. three-quarters of the total measures

i. half of the total measures

j. less than half of the total measures

Comments:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix C

Letter of Explanation for Piano Student and Parent

Université d'Ottawa · University of Ottawa

Faculté des arts Faculty of Arts
Musique Music

Catherine Lemay, M.A (Cand.)
Department of Music,
University of Ottawa
Ottawa, ON

Dear Piano Student and Parent,

My name is Catherine Lemay and I am presently involved in a research study conducted under the auspices of the University of Ottawa examining the validity and comparability of various sight-reading evaluation methods. The results of this study will form the basis of my Master's thesis at the University of Ottawa under the direction of Dr. Gilles Comeau in the music department.

As an active piano student between the ages of 12 and 15 and currently preparing for your grade seven examination as defined by the Royal Conservatory of Music your assistance in this project would be invaluable, and I invite you to participate in this study.

What you will be asked to do: Your participation will consist essentially of attending one short session at the Piano Pedagogy Research Laboratory at the University of Ottawa at which time you will be asked to sight-read approximately 12 short musical excerpts of various technical and musical difficulties. This session should take no longer than 25-30 minutes.

Where the study will be carried out: This study will take place at the Piano Pedagogy Research Laboratory, located in the Perez Hall, 50 Université, at the University of Ottawa.

Voluntary participation and anonymity of feedback: Your involvement in the study is voluntary, and at anytime you may choose not to participate. ***Additionally, at all times you will remain anonymous.*** The results of the study may be published, but your name will not appear on any of the documentation from the project. Your participation will bear no cost to you other than the time it takes to come to the Piano Pedagogy Research Laboratory and sight-reading short musical excerpts. The findings from this project will provide information on the validity and reliability of various sight-reading evaluation methods. No risks to participants beyond those experienced in everyday life are anticipated.

If you would like to participate in this project please contact Catherine Lemay at 704. If you have any questions about this research project, please feel free to contact.

Thank you for your time and consideration.
Sincerely,

Catherine Lemay

50 rue Université, P.O. Box 450 Stn. A
Ottawa, Ontario K1N 6N5 Canada
50 University St., PO Box 450 Stn. A
Ottawa Ontario K1N 6N5 Canada
Tel: (613) 562-5743 • Fax / Téléc: (613) 562-5140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix D

Parent Consent Form

Université d'Ottawa · University of Ottawa

Faculté des arts Musique | Faculty of Arts Music

Title of the Project: The Evaluation and Measurement of Music Reading Ability in Piano Students:
Comparing Assessment Systems

Catherine Lemay, M.A (Cand.) and Professor Gilles Comeau
Department of Music, University of Ottawa, Ottawa, ON

Invitation to Participate: My child is invited to participate in the research study mentioned above conducted by Catherine Lemay.

Purpose of the Study: The purpose of the study is to first adapt an existing sight-reading test to be used in the context of piano performance. This test will then be compared to two other systems of assessment, more specifically, a sight-reading error analysis and an evaluation by expert examiners, to determine whether these three systems of assessment provide similar evaluations of performances.

Participation: My child's participation will consist essentially of attending one session at the Piano Pedagogy Research Laboratory during which he/she will sight-read at the piano 12 musical excerpts of increasing difficulty. The session has been scheduled for _____ (place, date and time of the session).

Risks: My child's participation in this study will entail that he/she sight-reading musical material. His/her performance will be recorded by audio and by video camera and MIDI data, a standard system for recording and transferring musical information from the piano to a computer which makes possible the graphic representation of a musical performance.

Benefits: My child's participation in this study will help provide reliable criteria for the evaluation of sight-reading performance at the piano.

Confidentiality and anonymity: I have received assurance from the researcher that my child's sight-reading performances will remain strictly anonymous. I understand that the performance will be used only for the purpose of this research project and that his/her confidentiality will be protected by removing his/her name from all documentation. The results of the study may be published, but his/her name will not be linked to results in publications that are released from the project.

50, rue Université C.P. 450, Succ. A
Ottawa (Ontario) K1N 6N5 Canada

50 University St., P.O. Box 450, Stn. A
Ottawa, Ontario K1N 6N5 Canada

1 + (613) 562-5733 • Téléc./Fax (613) 562-5140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conservation of data: The information collected will consist of audio and video recordings as well as MIDI data. This data will be kept in a secure manner at the Piano Pedagogy Research Laboratory and will be available only to the researchers directly involved in this project.

Voluntary Participation: My child is under no obligation to participate and if he/she chooses to participate, he/she can withdraw from the study at any time. If my child chooses to withdraw, all data gathered until the time of withdrawal will be discarded and will not be used for the purpose of this project.

Acceptance: I, _____, agree to my child's participation in the research study conducted by Catherine Lemay of the Department of Music at the University of Ottawa, whose research is under the supervision of Dr. Gilles Comeau.

If I have any questions about the study, I may contact the researcher or her supervisor.

If I have any questions regarding the ethical conduct of this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5

Tel.: (613) 562-5841

Email: ethics@uottawa.ca

There are two copies of the consent form, one of which is mine to keep.

Parent's signature: Date:

Name of researcher:	Catherine Lemay
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music.
Phone number:	
Email address:	
Name of supervisor:	Gilles Comeau
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music
Phone number:	613.562.5800 ext. 3483
Email address:	gcomeau@uottawa.ca

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix E

Student Assent Form

Université d'Ottawa · University of Ottawa

Faculté des arts Musique | Faculty of Arts Music

Title of the Project: The Evaluation and Measurement of Music Reading Ability in Piano Students:
Comparing Assessment Systems

Catherine Lemay, M.A (Cand.) and Professor Gilles Comeau
Department of Music, University of Ottawa, Ottawa, ON

Invitation to Participate: I am invited to participate in the study mentioned above conducted by Catherine Lemay.

Purpose of the Study: The purpose of the study is to adapt an existing sight-reading test for wind instrumentalists to be used for pianists. This test will then be compared to the analysis of mistakes made while sight-reading and the judgment of experts.

Participation: My participation will consist of attending one session at the Piano Pedagogy Research Laboratory during which I will sight-read at the piano 12 short musical pieces of increasing difficulty. The session has been scheduled for _____ (place, date and time of the session).

Risks: My participation in this study will entail that I sight-read musical pieces. My performance will be recorded by audio, video camera and MIDI data, a system for recording and transferring musical information from the piano to a computer which makes possible the graphic representation of a musical performance.

Benefits: My participation in this study will help provide tools for the evaluation of sight-reading performance at the piano.

Confidentiality and anonymity: I have been assured by the researcher that my sight-reading performances will remain strictly anonymous. I understand that the performance will be used only for this research project and that my name will be removed from all documentation. The results of the study may be published, but my name will not be linked to results in publications that are released from the project.

Conservation of data: The information collected will consist of audio and video recordings as well as MIDI data. This data will be kept in a secure manner at the Piano Pedagogy Research Laboratory and will be available only to the researchers directly involved in this project.

50, rue Université C.P. 450, Succ. A
Ottawa (Ontario) K1N 6N5 Canada

50 University St., P.O. Box 450, Stn. A
Ottawa, Ontario K1N 6N5 Canada

1 + (613) 562-5733 • Téléc./Fax (613) 562-5140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time. If I choose to withdraw, all data gathered until the time of withdrawal will be thrown out and will not be used for the purpose of this project.

Acceptance: I, _____, agree to participate in the above research study conducted by Catherine Lemay of the Department of Music at the University of Ottawa, whose research is under the supervision of Dr. Gilles Comeau.

If I have any questions about the study, I may contact the researcher or her supervisor.

If I have any questions regarding the ethical conduct of this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5

Tel.: (613) 562-5841

Email: ethics@uottawa.ca

There are two copies of the consent form, one of which is mine to keep.

Participant's signature: Date:

Name of researcher:	Catherine Lemay
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music.
Phone number:	
Email address:	
Name of supervisor:	Gilles Comeau
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music
Phone number:	613.562.5800 ext. 3483
Email address:	gcomeau@uottawa.ca

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix F

Student Form

Name: _____

Gender: Male / Female

1. How old are you? _____
2. How many years have you been taking piano lessons? _____
3. List all piano methods with which you have learnt? _____

   _____
4. What is your level? _____
5. How many days per week do you practice your piano? _____
   a. How long do your practice sessions last? _____
   b. Does someone help you with your piano practice? _____
6. How often do you sight-read music?
   a. Every day
   b. Every few days
   c. Once a week
   d. A few times per month
   e. Once a month
   f. Almost never
7. Do you enjoy sight-reading music? _____
8. When you compare your sight-reading abilities with those of other students, do you consider yourself:
   a. Better than most students
   b. Better than average
   c. About average
   d. Worse than average
   e. Worse than most students

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix G

Exercise 1

***Exercises used for this research project have been removed from this thesis document as they will be converted into a Sight-reading Scale for the piano. For further information contact the Piano Pedagogy Research Laboratory at the University of Ottawa.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix H

Exercise 2

***Exercises used for this research project have been removed from this thesis document as they will be converted into a Sight-reading Scale for the piano. For further information contact the Piano Pedagogy Research Laboratory at the University of Ottawa.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix I

Exercise 3

***Exercises used for this research project have been removed from this thesis document as they will be converted into a Sight-reading Scale for the piano. For further information contact the Piano Pedagogy Research Laboratory at the University of Ottawa.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix J

Exercise 4

***Exercises used for this research project have been removed from this thesis document as they will be converted into a Sight-reading Scale for the piano. For further information contact the Piano Pedagogy Research Laboratory at the University of Ottawa.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix K

Exercise 5

***Exercises used for this research project have been removed from this thesis document as they will be converted into a Sight-reading Scale for the piano. For further information contact the Piano Pedagogy Research Laboratory at the University of Ottawa.**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix L

Letter of Explanation for Graduate Students

Université d'Ottawa · University of Ottawa

Catherine Lemay, M.A (Cand.)
Department of Music,
University of Ottawa
Ottawa, ON

Dear Graduate Student,

My name is Catherine Lemay and I am presently involved in a research study conducted under the auspices of the University of Ottawa examining the validity and comparability of various sight-reading evaluation methods. The results of this study will form the basis of my Master's thesis at the University of Ottawa under the direction of Dr. Gilles Comeau in the music department. As graduate student your assistance in this project would be invaluable, and I invite you to participate in this study.

What you will be asked to do: Your participation will consist essentially of evaluating the audio recording of the sight-reading performance of 8 young piano students with the use of an evaluation grading chart which will be given to you at the time of your session.

Voluntary participation and anonymity of feedback: Your involvement in the study is voluntary, and at anytime you may choose not to participate. ***Additionally, at all times you will remain anonymous.*** The results of the study may be published, but your name will not appear on any of the documentation from the project. Your participation will bear no cost to you other than the time it takes to evaluate the performances. The findings from this project will provide information on the validity and reliability of various sight-reading evaluation methods. No risks to participants beyond those experienced in everyday life are anticipated.

If you would like to participate in this project please contact Catherine Lemay at . If you have any questions about this research project, please feel free to contact me at any time.

Thank you for your time and consideration.
Sincerely,

Catherine Lemay

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix M

Graduate Student Consent Form

Université d'Ottawa · University of Ottawa

Faculté des arts Musique | Faculty of Arts Music

Title of the Project: The Evaluation and Measurement of Music Reading Ability in Piano Students: Comparing Assessment Systems

Catherine Lemay, M.A (Cand.) and Professor Gilles Comeau
Department of Music, University of Ottawa, Ottawa, ON

Invitation to Participate: You are invited to participate in the research study mentioned above conducted by Catherine Lemay.

Purpose of the Study: The purpose of the study is to first adapt an existing sight-reading test to be used in the context of piano performance. This test will then be compared to two other systems of assessment, more specifically, a sight-reading error analysis and an evaluation by expert examiners, to determine whether these three systems of assessment provide similar evaluations of performances.

Participation: Your participation will consist essentially of attending one session at the Piano Pedagogy Research Laboratory during which you will evaluate the audio recording of the sight-reading performance of 10 young piano students with the use of an evaluation grading chart which will be given to you at the time of your session. Your session has been scheduled for _____ (place, date and time of the session).

Benefits: My participation in this study will help provide reliable criteria for the evaluation of sight-reading performance at the piano.

Confidentiality and anonymity: Your participation in this research project will remain strictly anonymous. Your evaluation will be used only for the purpose of this research project and that your confidentiality will be protected by removing your name from all documentation. The results of the study may be published, but your name will not be linked to results in publications that are released from the project.

Conservation of data: The information collected will consist of written evaluations according to evaluation guidelines. This data will be kept in a secure manner at the Piano Pedagogy Research Laboratory and will be available only to the researchers directly involved in this project.

50, rue Université C.P. 450, Succ. A
Ottawa (Ontario) K1N 6N5 Canada

50 University St., P.O. Box 450, Stn. A
Ottawa, Ontario K1N 6N5 Canada

1 + (613) 562-5733 • Telec./Fax (613) 562-5140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Voluntary Participation: You are under no obligation to participate and if you choose to participate, you can withdraw from the study at any time. If you choose to withdraw, all data gathered until the time of withdrawal will be discarded and will not be used for the purpose of this project.

Acceptance: I, _____, agree to participate in the above research study conducted by Catherine Lemay of the Department of Music at the University of Ottawa, whose research is under the supervision of Dr. Gilles Comeau.

If you have any questions about the study, you may contact the researcher or her supervisor.

If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5

Tel.: (613) 562-5841

Email: ethics@uottawa.ca

There are two copies of the consent form, one of which is yours to keep.

Participant's signature: Date:

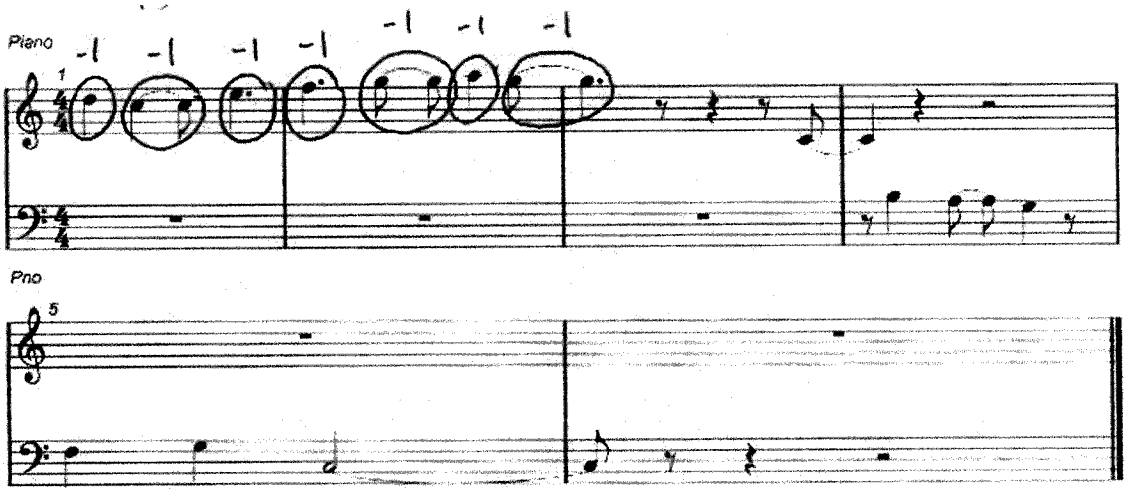
Name of researcher:	Catherine Lemay
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music.
Phone number:	
Email address:	
Name of supervisor:	Gilles Comeau
Institution, Faculty, Department:	University of Ottawa, Faculty of Arts, Department of Music
Phone number:	613.562.5800 ext. 3483
Email address:	gcomeau@uottawa.ca

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

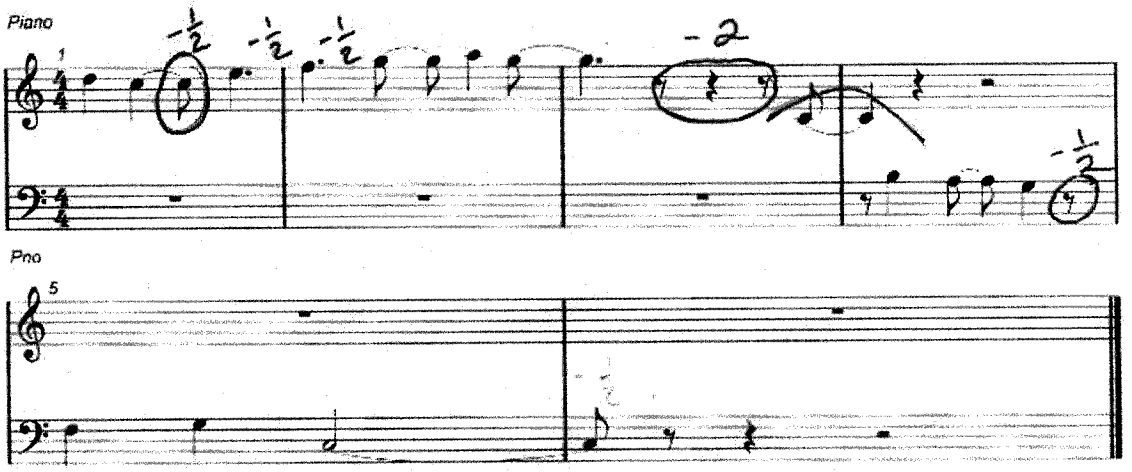
Appendix N

Quantification of Exercise 1 Subject D according to the GSA

Pitch



Rhythm

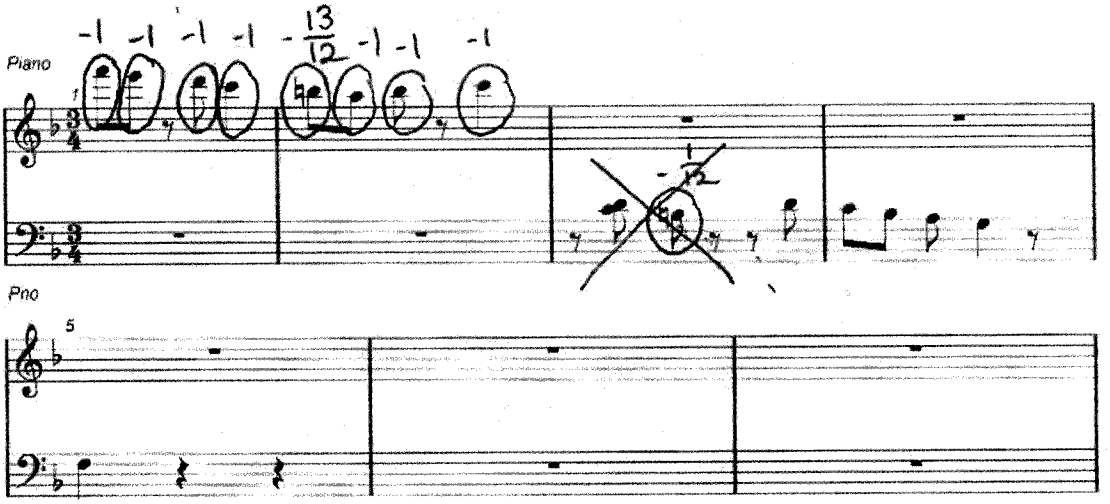


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

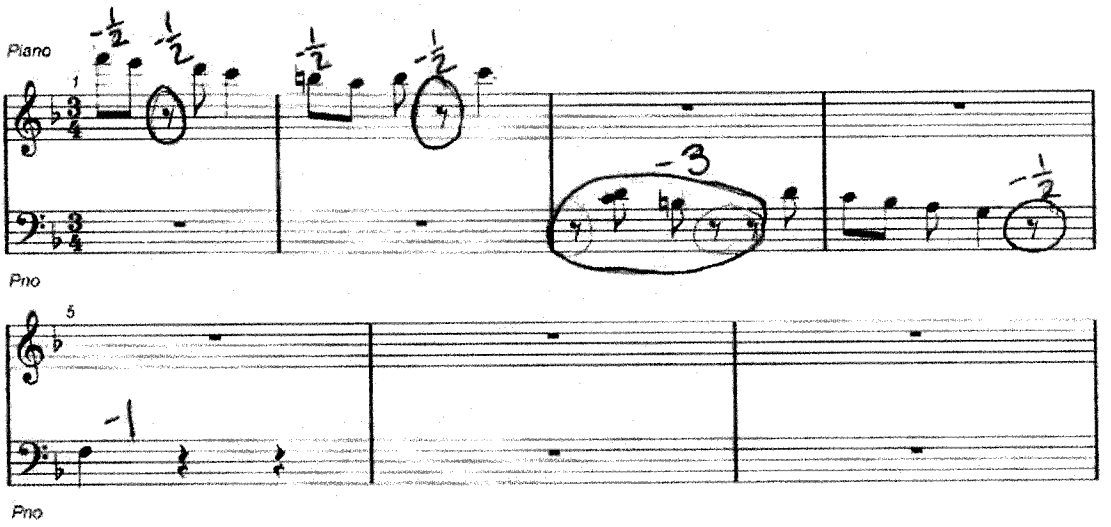
Appendix O

Quantification of Exercise 2 Subject D according to the GSA

Pitch



Rhythm

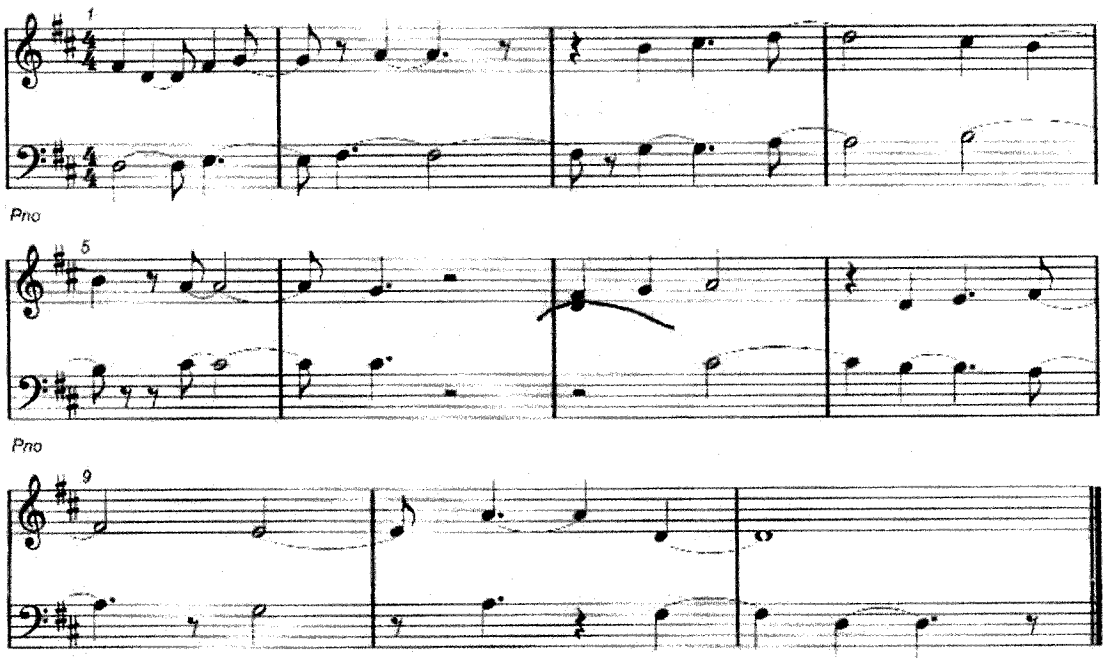


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

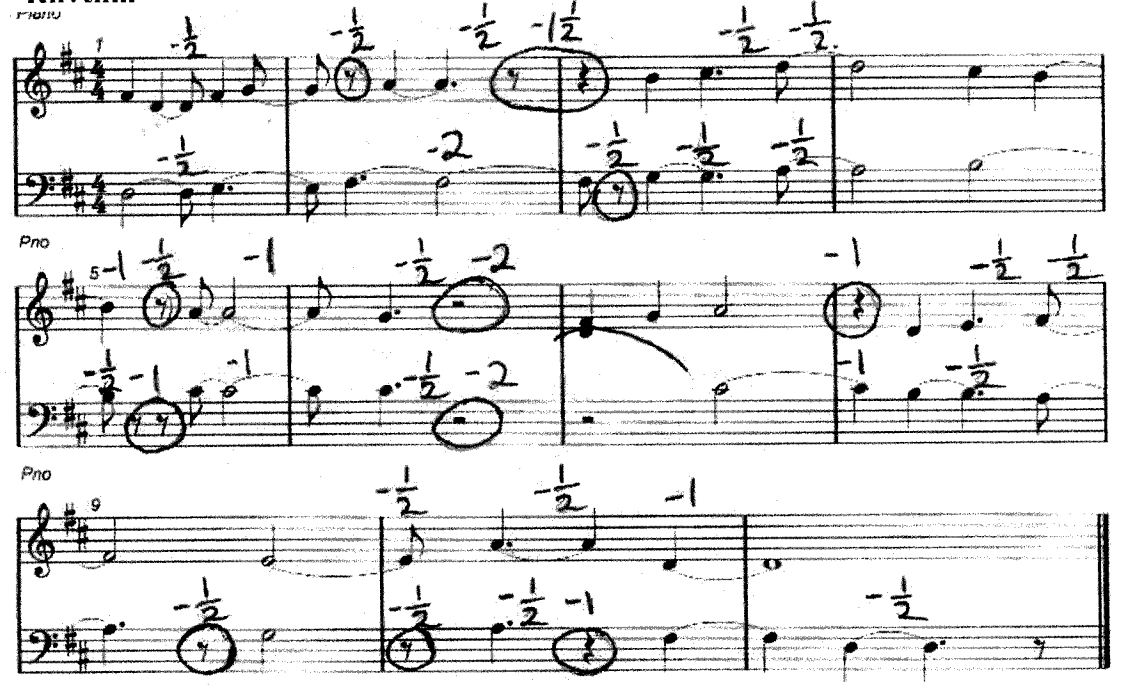
Appendix P

Quantification of Exercise 3 Subject D according to the GSA

Pitch



Rhythm

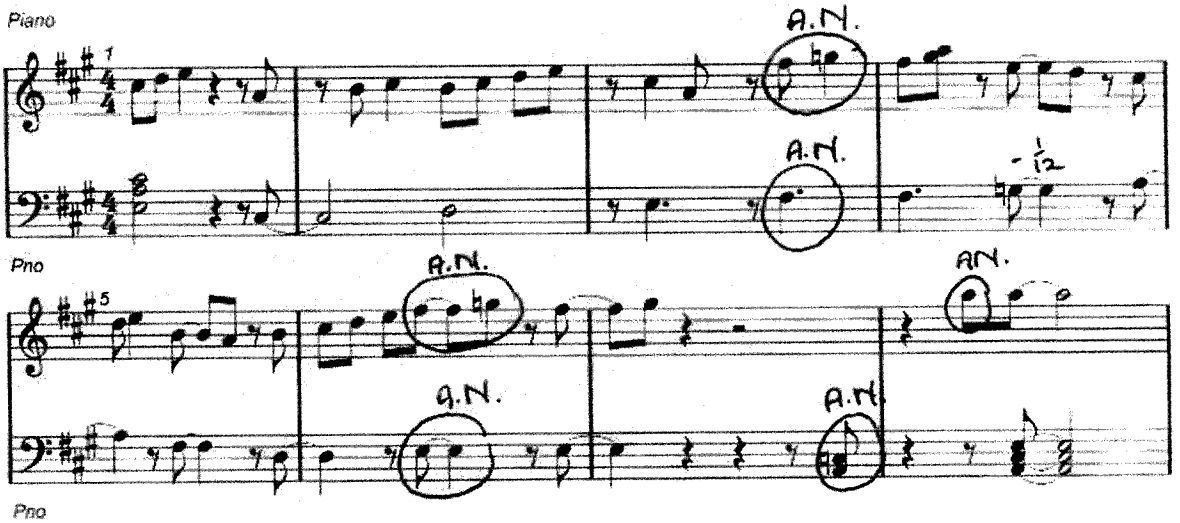


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

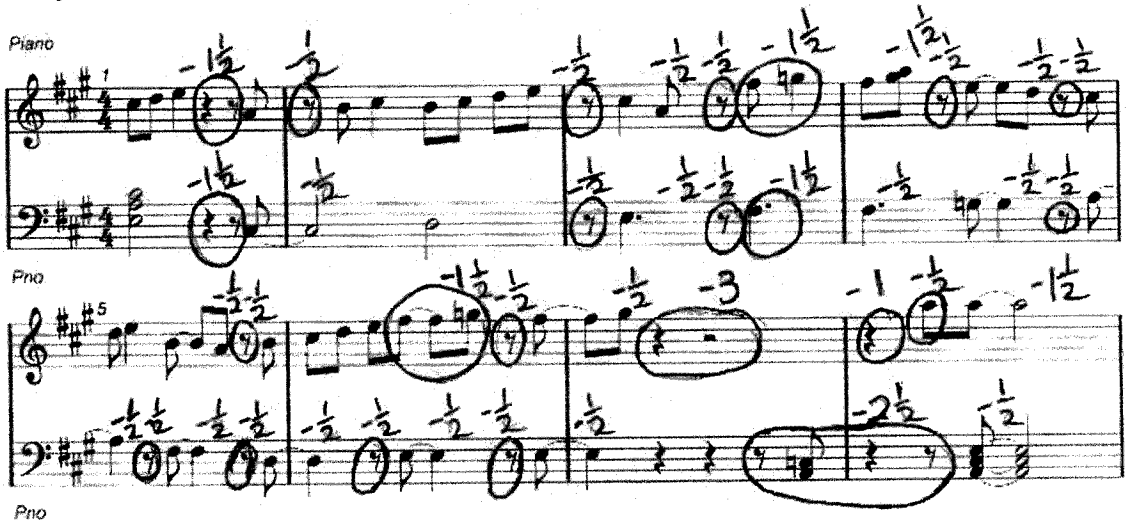
Appendix Q

Quantification of Exercise 4 Subject D according to the GSA

Pitch



Rhythm

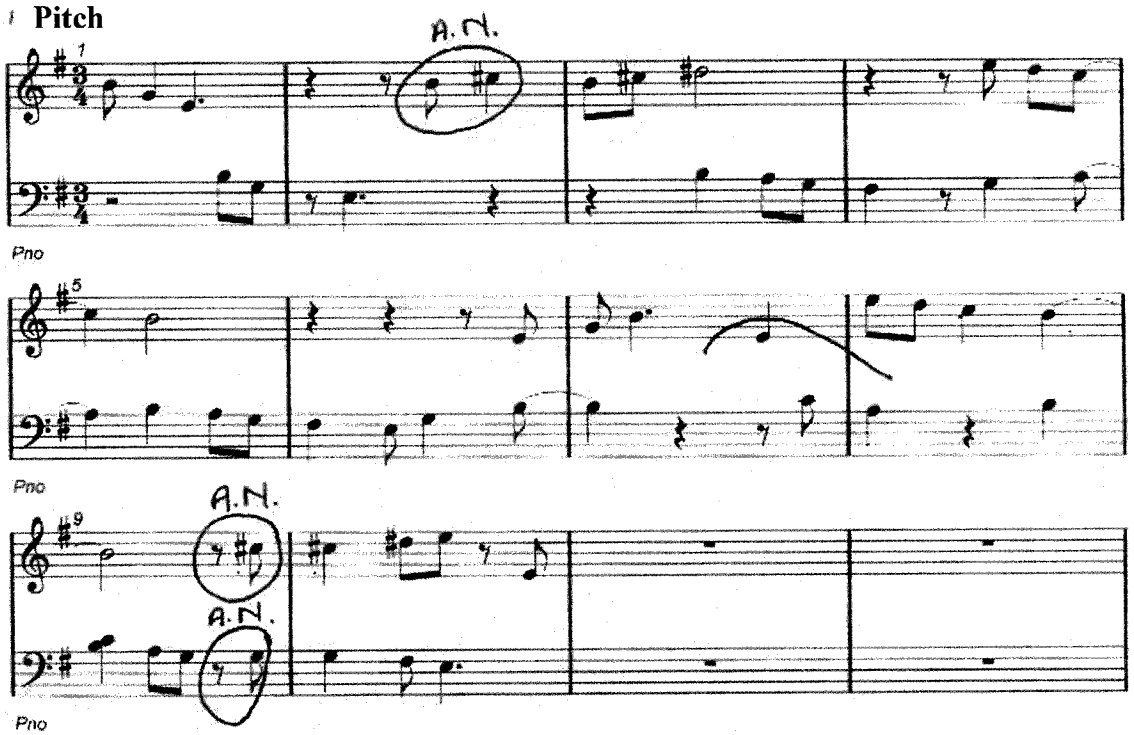


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

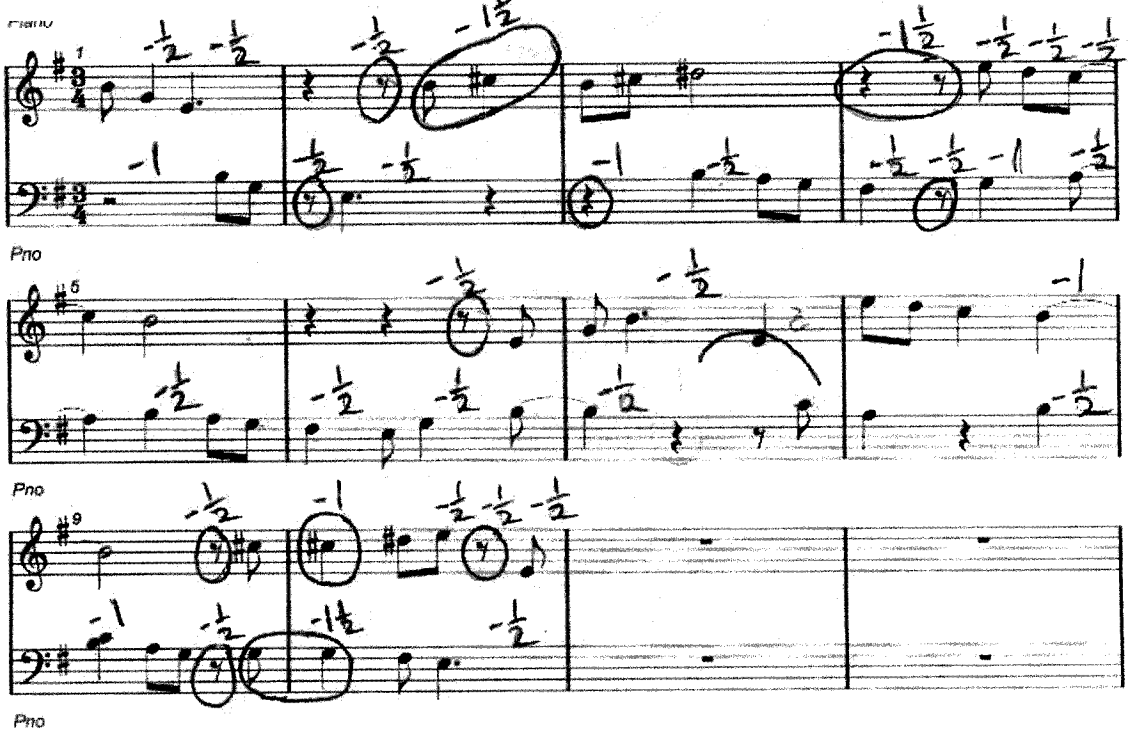
Appendix R

Quantification of Exercise 5 Subject D according to the GSA

Pitch



Rhythm



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix S

Examiner Consent Form

Université d'Ottawa · University of Ottawa

Faculté des arts Musique | Faculty of Arts Music

Title of the Project: The Evaluation and Measurement of Music Reading Ability in Piano Students: Comparing Assessment Systems

Catherine Lemay, M.A (Cand.) and Professor Gilles Comeau
Department of Music, University of Ottawa, Ottawa, ON

Invitation to Participate: You are invited to participate in the research study mentioned above conducted by Catherine Lemay.

Purpose of the Study: The purpose of the study is to first adapt an existing sight-reading test to be used in the context of piano performance. This test will then be compared to two other systems of assessment, more specifically, a sight-reading error analysis and an evaluation by expert examiners, to determine whether these three systems of assessment provide similar evaluations of performances.

Participation: Your participation will consist essentially of adjudicating the audio recording of the sight-reading performance of eight young piano students and provide comments on your adjudication criteria and procedure.

Benefits: My participation in this study will help provide reliable criteria for the evaluation of sight-reading performance at the piano.

Confidentiality and anonymity: Your participation in this research project will remain strictly anonymous. Your adjudication and comments will be used only for the purpose of this research project and that your confidentiality will be protected by removing your name from all documentation. The results of the study may be published, but your name will not be linked to results in publications that are released from the project.

Conservation of data: The information collected will consist of written adjudication and comments. This data will be kept in a secure manner at the Piano Pedagogy Research Laboratory and will be available only to the researchers directly involved in this project.

50, rue Université C.P. 450, Succ. A
Ottawa (Ontario) K1N 6N5 Canada

50 University St., P.O. Box 450, Stn. A
Ottawa, Ontario K1N 6N5 Canada

1 + (613) 562-5733 • Teléc./Fax (613) 562-5140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Voluntary Participation: You are under no obligation to participate and if you choose to participate, you can withdraw from the study at any time. If you choose to withdraw, all data gathered until the time of withdrawal will be discarded and will not be used for the purpose of this project.

Acceptance: I, _____, agree to participate in the above research study conducted by Catherine Lemay of the Department of Music at the University of Ottawa, whose research is under the supervision of Dr. Gilles Comeau.

If you have any questions about the study, you may contact the researcher or her supervisor.

If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5

Tel.: (613) 562-5841

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix O

Examiner Scoring Sheets

Instructions for grading:

1. Please listen to each audio clip provided:
    a. Each clip includes one piano student's sight-reading performance of a short exercise.
    b. Each student performs five exercises.
    c. Therefore, clips are labeled a1, a2, a3, a4, a5, for the first subject, then b1, b2... for the second subject, and so forth.
    d. The sample includes a total of 8 piano students.

2. After listening to each clip, please provide a grade out of ten for that particular performance. (Ex. a1. ? /10) Grades can have half points (Ex. a1. 8.5/10).

3. For this project, we will go by the RCM standard and use 6/10 to indicate a passing grade. Any mark below 6 is a failing grade.

4. For each clip, please provide comments on your grading strategies as well as your reasoning behind providing each grade. Examples:

Ex. 6/10 - **Grading strategies:** This performance was extremely unsteady rhythmically. Some misconception of 8ths at the beginning of m. 2. A hesitant and incomplete performance (last note missing). Nevertheless, the candidate played the correct pitches of the notes. **Reasoning:** At this level I would pass the candidate, but not give a higher mark.

Ex. 7/10 - **Grading strategies:** This performance was above average, but not in the first class honors category. I had the feeling that the candidate was not counting from the outset. This is a situation where the player understands note values, but not within their context. The student needs to understand the importance of meter and consistency of pulse. **Reasoning:** Dynamics were not observed; in fact the student played the opposite *f* became *p* at m. 5 Poor sense of rhythmic structure.

Ex. 8.5/10 - **Grading strategies:** Definitely a first class honors reading. But this candidate stopped to correct notes twice, B flat to B, C natural to C#. Otherwise the playing was musical and fluent. **Reasoning:** Because of the stops for corrections, I did not award 90. The playing is outstanding really.

Ex. 10/10 - **Grading strategies:** This was an excellent performance, with a quick uninterrupted tempo and musical shaping of the phrase from the first note. Fine legato playing. **Reasoning:** This was clearly a more advanced student reading this piece. It's beyond what most Grade 1 level students could accomplish.

5. After listening to all clips from each subject, please calculate the subject's final score for a total grade out of fifty.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6. There is one grading sheet for each subject.

7. Once you have listened to each subject's performance, the final step is to rate these subjects according to their overall performance on the five exercises.

Thank you very much for your time as well as sharing your expertise by participating in this research project.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject a
Total grade: /50

a1. /10

Grading strategies:

Reasoning:

---

a2. /10

Grading strategies:

Reasoning:

---

a3. /10

Grading strategies:

Reasoning:

---

a4. /10

Grading strategies:

Reasoning:

---

a5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject b
Total grade: /50

b1. /10

Grading strategies:

Reasoning:

---

b2. /10

Grading strategies:

Reasoning:

---

b3. /10

Grading strategies:

Reasoning:

---

b4. /10

Grading strategies:

Reasoning:

---

b5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject c
Total grade: /50

c1. /10

Grading strategies:

Reasoning:

---

c2. /10

Grading strategies:

Reasoning:

---

c3. /10

Grading strategies:

Reasoning:

---

c4. /10

Grading strategies:

Reasoning:

---

c5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject d
Total grade: /50

d1. /10

Grading strategies:

Reasoning:

---

d2. /10

Grading strategies:

Reasoning:

---

d3. /10

Grading strategies:

Reasoning:

---

d4. /10

Grading strategies:

Reasoning:

---

d5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject e
Total grade: /50

e1. /10

Grading strategies:

Reasoning:

---

e2. /10

Grading strategies:

Reasoning:

---

e3. /10

Grading strategies:

Reasoning:

---

e4. /10

Grading strategies:

Reasoning:

---

e5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject f
Total grade: /50

f1. /10

Grading strategies:

Reasoning:

---

f2. /10

Grading strategies:

Reasoning:

---

f3. /10

Grading strategies:

Reasoning:

---

f4. /10

Grading strategies:

Reasoning:

---

f5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject g
Total grade: /50

g1. /10

Grading strategies:

Reasoning:

---

g2. /10

Grading strategies:

Reasoning:

---

g3. /10

Grading strategies:

Reasoning:

---

g4. /10

Grading strategies:

Reasoning:

---

g5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subject h
Total grade: /50

h1. /10

Grading strategies:

Reasoning:

---

h2. /10

Grading strategies:

Reasoning:

---

h3. /10

Grading strategies:

Reasoning:

---

h4. /10

Grading strategies:

Reasoning:

---

h5. /10

Grading strategies:

Reasoning:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RANKING OVERALL PERFORMANCE

The final step of the evaluation is to rate these subjects according to their overall performance on the five exercises. The subject in first place being the best performer and the one in eighth place being the worst. (Ex. First place: subject "a")

First place:

Second place:

Third place:

Fourth place:

Fifth place:

Sixth place:

Seventh place:

Eighth place:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix P

Grade Reductions with Each Method of Assessment for Subject A

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	50%	0%	25% Articulation	2.68%	28.57%	23.33% Mainly for pitch
Exercise 2	0%	75%	Articulation	0.00%	26.79%	35% Mainly for rhythm
Exercise 3	62.5%	25%	Articulation	1.69%	25.68%	31.67% Mainly for pitch
Exercise 4	100%		Articulation	1.03%	32.02%	46.67% For both pitch and rhythm errors
Exercise 5	25%	37.5%	12.5% Articulation	1.15%	25.49%	46.67% Mainly for rhythm

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix Q

Grade Reductions with Each Method of Assessment for subject B

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	0%	0%	none	0%	10.71%	0%
Exercise 2	0%	0%	25% Articulation	0%	23.21%	1.67% Articulation
Exercise 3	0%	12.5%	none	0%	18.92%	1.67% Rhythm
Exercise 4	16.67%	16.67%	16.67% Expression	0%	8.99%	11.67% Pitch and expression
Exercise 5	0%	43.75%	none	0.41%	17.16%	16.67% Pitch, expression, and timing

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix R

Grade Reductions with Each Method of Assessment for Subject C

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	25%	25%	12.5%	2.39%	33.93%	21.67% Pitch and Timing
Exercise 2	87.5%		None	5.96%	26.79%	43.33% Pitch, Rhythm, and Timing
Exercise 3	75%	12.5%	None	2.59%	114.86%	53.33% Pitch, Rhythm, and Timing
Exercise 4	83.33%	8.33%	None	3.93%	34.27%	51.67% Pitch, Timing, Rhythm, and Expression
Exercise 5	100%		None	1.15%	126.96%	51.67% Pitch, Timing, and Pulse

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix S

Grade Reductions with Each Method of Assessment for Subject D

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	50%	0%	25% Articulation	25%	16.07%	16.67 Pitch
Exercise 2	87.5%		None	27.96%	23.21%	41.67 Pitch and Timing
Exercise 3	0%	43.75% Rhythm and Articulation		0%	22.3%	16.67% Timing
Exercise 4	50%	0%	41.67% Articulation	0.09%	35.39%	28.33% Pitch, Timing, and Expression
Exercise 5	12.5%	62.5%	6.26% Timing	0.98%	22.55%	33.33% Timing and Rhythm

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix T

Grade Reductions with Each Method of Assessment for Subject E

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	0%	25%	50% Articulation	0%	28.57%	8.33% Timing
Exercise 2	50%	0%	25% Articulation	0.61%	14.29%	21.67 Pitch and Tempo
Exercise 3	12.5%	31.25%	25% Timing and Articulation	0%	48.65%	22.33% Rhythm and Timing
Exercise 4	50%	33.33%	Articulation and Expression	0%	48.31%	33.33% Pitch, Rhythm, and Timing
Exercise 5	0%	62.5%	6.25% Expression Articulation	0%	22.06%	33.33% Rhythm, Timing, and Expression

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix U

Grade Reductions with Each Method of Assessment for Subject F

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	0%	0%	50% Articulation	0%	10,71%	0%
Exercise 2	0%	0%	75% Articulation	0%	14,29%	16.67% Articulation and Tempo
Exercise 3	0%	0%	62.5% Articulation	0%	22,97%	33.33% Articulation
Exercise 4	33.33%	0%	50% Articulation and Expression	0%	10.11%	10% Pitch and Expression
Exercise 5	12.5%	56.25%	12.5% Timing	0%	58,33%	15% Expression, Timing, and Tempo

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix V

Grade Reductions with Each Method of Assessment for Subject G

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	0%	0%	62.5% Articulation	0%	16.07%	0%
Exercise 2	25%	50%	25% Articulation	3.57%	17.86%	31.67% Timing, Rhythm, Missing note
Exercise 3	93.75% Pitch, Rhythm, and Articulation			1.46%	55.41%	30.5% Tempo and Rhythm
Exercise 4	16.67%	83.33%	Articulation	1.12%	61.8%	46.67% Timing, Rhythm, and Pitch
Exercise 5	12.5%	75% Rhythm and Articulation		0.41%	54.41%	45% Timing, Rhythm, and Pitch

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix W

Score Reductions with Each Method of Assessment for Subject H

	PSS			GSA		EE
	Pitch	**Rhythm**	**Other**	**Pitch**	**Rhythm**	
Exercise 1	0%	0%	50% Articulation	0%	3.57%	0%
Exercise 2	0%	75% Rhythm and Articulation		0%	19.64%	21.67% Rhythm and Timing
Exercise 3	25%	0%	43.75% Articulation and Timing	0%	21.62%	20% Pitch
Exercise 4	100% Pitch, Rhythm, and Articulation			1.13%	73.6%	30% Pitch, Tempo, and Rhythm
Exercise 5	25%	37.5%	12.5% Timing	0.17%	18.14%	16.67% Expression and Pitch

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.