THE EFFECT OF VIDEOTAPE ANALYSIS AND PEER CRITIQUE ON DEVELOPING THE SKILL OF SELF-ASSESSMENT IN PRE-SERVICE TEACHERS

by

DONALD MORAN SKAGGS III

(Under the Direction of Mary Leglar)

ABSTRACT

Writing on teacher evaluation in the *Handbook of Research on Music Teaching and Learning* (1992), Taebel aptly stated that: "Teacher evaluation is a hazardous and complex undertaking, perhaps because the concepts of teaching and evaluation are multifaceted and complex" (p. 310). Because of the complexity of the task, many approaches to assessing the effectiveness of teaching have been explored. Three of these command most attention in the literature: self-assessment, peer critique, and supervisor evaluation. A definitive answer to which of these three techniques yields best results continues to claim the attention of the research community.

The purpose of this study was to determine the effect of videotape analysis and peer critique on developing the skill of self-assessment in pre-service teachers. Four research questions were posed: 1) Does the use of videotape analysis and a peer critique process result in more accurate self-assessment? 2) Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate? 3) To what extent do the self-assessments agree with those

of experts? 4) Do identifiable personality traits affect accuracy in self-assessment: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience? The study employed a pretest posttest design. Participants (N=40) were undergraduate students enrolled in either an elementary music methods course for music education majors or a music methods course for non-music education majors. The subjects were randomly divided into a Control Group and three Experimental Groups. Before beginning the study, participants completed the *Internet Personality Inventory Survey* designed to measure each of five personality domains: extraversion, agreeableness, conscientiousness, emotional stability, and openness.

Each participant was required to teach four lessons. The Control Group completed a self-assessment rubric immediately following each lesson without receiving any exterior input. Experimental Group I completed the rubric in like fashion and then sought validation by viewing a video recording; Group II completed the rubric and then sought validation by reviewing peer critiques; Group III completed the assessment rubric and validated with both videotape analysis and peer critiques.

Quantitative measures were employed to analyze the data for questions one, three, and four; qualitative measures were used to analyze the data for question two. Findings revealed no significant difference in the self-assessment scores of the four groups, indicating that the use of videotape analysis and peer critique did not improve ability to self-evaluate. Although there were no significant differences, data revealed that participants using videotape analysis were able to identify teaching weaknesses at a quicker rate. Peer critiques were often found to be inaccurate and not time efficient. No significant difference in four of the five personality traits was found between groups. A

significant difference was found between the Control Group and Experimental Group III for the quality of “openness.” Based on the data collected in this study, it appears that personality traits are not an influential factor in the process of self-assessment.

INDEX WORDS: Self-assessment, Peer critique, Videotape Analysis, Teacher Training, Teacher Education

THE EFFECT OF VIDEOTAPE ANALYSIS AND PEER CRITIQUE ON DEVELOPING THE SKILL OF SELF-ASSESSMENT IN PRE-SERVICE TEACHERS

by

DONALD MORAN SKAGGS III

B.M,E., University of Southern Mississippi, 2001

M.M.E., University of Kansas, 2003

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF EDUCATION

ATHENS, GEORGIA

2009

THE EFFECT OF VIDEOTAPE ANALYSIS AND PEER CRITIQUE ON DEVELOPING THE SKILL OF SELF-ASSESSMENT IN PRE-SERVICE TEACHERS

by

DONALD MORAN SKAGGS III

Major Professor: Mary Leglar

Committee: Roy Kennedy
Michael Robinson
Clinton Taylor
Stephen Valdez

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

DEDICATION

This project is dedicated to my Lord and Savior Jesus Christ and to my family. First, to my Lord and Savior who has allowed me to finish this degree and to pursue a calling to teach. Without his guidance and direction I would have never made it to this point. Second, to my parents, (by birth) Don and Shirley Skaggs and (by marriage) Russ and Carol Shinpoch, whose nurturing and unconditional love gave me the strength to pursue my educational and life goals; to my daughter Anna Kate, who has brought untold joy to my life and put up with her daddy being away from her each week; and finally, to my wife, Joanne, whose love has endured, whose support has always been available, and who has always encouraged, prayed for, comforted, and inspired.

ACKNOWLEDGEMENTS

This project would not have been possible without the support and encouragement of the following people. First, my committee: Dr. Roy Kennedy, Dr. Donald Lowe, Dr. Michael Robinson, Dr. Clinton Taylor, Dr. Stephen Valdez, and Dr. William Wraga. Thank you for working with me and guiding me through the process of studying for my comprehensive exams, selecting my dissertation topic, defending the prospectus, the public presentation, and the final defense. Second, my chair and mentor Dr. Mary Leglar, for setting high expectations for me. Thank you for teaching me perseverance and a love of knowledge. You pushed me further than I ever thought possible and for that I am forever grateful. You are an incredible teacher and I am so glad that I was able to have three years to learn from you. Third, Dr. Roy Legette, Mrs. Carolyn Alford, and Ms. Zandra Bell for allowing me to conduct my research in their classes. Fourth, to the students of elementary methods and elementary music for the non-music education major for volunteering your time to help with this study. Finally, to Mr. Brandon Robertson for his help with data collection and evaluation of the participants.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Writing on teacher evaluation in the *Handbook of Research on Music Teaching and Learning*, Taebel aptly stated: "Teacher evaluation is a hazardous and complex undertaking, perhaps because the concepts of teaching and evaluation are multifaceted and complex" (p. 310). Because of the complexity of the task, many approaches to assessing the effectiveness of teaching have been explored. Three of these command most attention in the literature: self-assessment, peer critique, and supervisor evaluation. A definitive answer to which of these three techniques yields best results continues to claim the attention of the research community.

Self-assessment suggests a continual examination of one's teaching practices and techniques, and can employ a variety of methods in the process, including audio recording, video recording, journaling, and portfolios, to name a few. Of these methods, the most widely researched is the use of video. The use of video allows the teacher to review a given instructional segment without relying on recall, thus lending the feature of authenticity. It also allows for repetitive viewing, an asset for developing the skill of reflection.

Peer critique traditionally draws on the expertise of one or more individuals having professional experience the same as, or equal to, that of the person being reviewed. In practice peer evaluators work closely with the teacher on all aspects of instruction: instructional planning, selection of teaching strategies, and critical

observation of the classroom. Peer critique should typically function as a non-threatening collaborative approach to improving teaching.

Supervisor evaluation is completed by any number of individuals serving as overseers of a group, in this case, teachers. The supervisor usually completes the evaluation while the teaching episode is occurring. Similar to peer critique, the supervisor's role should be non-threatening. An important component of supervisor evaluation is the pre-observation and post-observation conference. Ideally, during the evaluation process the supervisor should, through questioning, lead the teacher to "self-realize" potential areas for improvement. It is at this juncture of the process that strategies, techniques, and solutions are best suggested (Zepeda, 2007).

Need

Although there is a significant amount of literature relating to evaluation, much of it has resulted in either conflicting or inconclusive results. Research studies have revealed minimal evidence of the effects of peer critique and videotape analysis on an individual's ability to self-evaluate. This study will seek to explore these effects and address these concerns.

Purpose of the Study

The purpose of this study was to determine the effect of videotape analysis and peer critique on developing the skill of self-assessment in pre-service teachers. This study sought to identify: a) the changes in students' self-assessments over the course of the study, b) the influence of videotape analysis and/or peer critique on the participant's self-assessment, and c) the degree of difference between participants who received no

videotape analysis or peer critique and those participants who did receive either videotape analysis, peer critique, or a combination of both videotape analysis and peer critique.

Research Questions

Four research questions guided the study:

1) Does the use of videotape analysis and a peer critique process result in more accurate self-assessment?

2) Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate?

3) To what extent do the self-assessments agree with those of experts?

4) Do identifiable personality traits affect accuracy in self-assessment: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience?

Definitions

Definitions for the following terms are provided to aid the reader in understanding the key premises on which this study is based.

Preservice Teacher: an undergraduate general education or music education major that was enrolled in a music education course. Most of these individuals were in their junior year of college and would be student teaching within one year.

Self-assessment: a reflection of one's own teaching using one of four forms (Appendices D-G). Self-assessment was completed within 24 hours of the teaching episode.

Peer Critique: an analysis of a peer's teaching utilizing the appropriate lessons forms (Appendices D-G). Peer critiques were completed immediately following the teaching episode.

Video recording: a stationary camera placed in the back of the room and set to wide angle. The teaching episode was captured in its entirety. The participant used the video recording as an aid in completing the self-assessment.

Assumptions

This study assumed that students have had some experience in the processes of self-assessment. Although the depth and breadth of understanding varied from subject to subject, all participants should have been prepared to complete the prescribed evaluation forms.

A second assumption is that the peer critiques were objective. Since the evaluation forms remained confidential, participants completing peer critiques should have been able to complete the prescribed evaluation forms with minimal bias.

A third assumption is that the committee of professional evaluators was accurate. The study assumed that their evaluations of the participants teaching would match those of other professionals in the field of music education.

The fourth assumption is that participants answered the questions to the personality test as accurately and honestly as possible. The study assumes that students took the time to think about and correctly mark each response.

Limitations

The scope of this study was limited to exploring the effect of videotape analysis and peer critique on developing the skill of self-assessment in pre-service teachers.

Participants were undergraduate preservice teachers enrolled in either an elementary music methods course or a music education for non-music majors course. Due to the size of the sample, geographical location, and the contextual setting, any attempts to generalize the findings of this study should be approached with caution.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

In an era where educational demands are high, educational systems are focusing on everything from high stakes testing to a thorough evaluation of the curriculum, students, and teachers. Interfering with these focuses are constant technological change and social mobility, which present many challenges for the practicing teacher (Kremer-Hayon, 1993). A common goal of educational institutions has become continuous improvement while accommodating these changes. Two views on how this change is made exist. The first is that change results from a top-down approach where the leader makes all decisions and guides the group while the second is that change results from a bottom-up approach. This approach places greater focus on the individual classroom teacher.

Acknowledging the importance of the individual teacher in the bottom-up approach, the question becomes: How can an institution make sure its teachers are functioning at their highest level? The answer: through teacher evaluation. As Stronge and Tucker (2003) state: "Without capable, high quality teachers in America's classrooms, no educational reform effort can possibly succeed" (p. 3). This brings about several other questions: What is evaluation? What are the types of evaluation? What is the benefit of evaluation? How does one go about evaluating? Is there one evaluation

technique that is better than another? How does evaluation affect teacher training programs?

Evaluation defined is “to determine the significance, worth, or condition of a trait, usually by careful appraisal and study” (Webster, 1953, p. 834). It allows individuals to select choices and come to decisions (Simpson, 1966). There are several types of teacher evaluation. These include self-assessment, peer critique, and supervisor evaluation. Self-assessment includes examining or appraising an aspect of oneself. Dewey promotes this reflective thinking because it “converts action that is merely appetitive, blind, and impulsive into intelligent action” (Dewey, 1933, p. 17). The evaluation can be completed using a variety of methods (Roth & Tobin, 2000). These include: videotaping, audio taping, reflective dialogue, self-reports, portfolios, and journal writing (Scriven et al., 1990). Peer critique involves evaluating or appraising another individual who is at a similar level as that of the evaluator. Peer critique can be completed through live observation, prerecorded observation, discussion groups, or collaborative learning. Supervisor evaluation refers to a non-threatening appraisal by an individual considered to be in a higher-level position than that of the person being evaluated. This individual can range from an administrator or lead teacher in the public school to a professor or graduate teaching assistant at the higher education level.

In order to gain a clearer understanding of the problems posed in this study it is imperative that prior evaluation and assessment research be thoroughly examined. Changes in self-assessment, peer critique, and supervisor evaluation were explored followed by general assessment and teacher training issues.

Self-assessment

Self-assessment is a process that allows individuals to improve or better themselves by examining and critiquing an aspect or several aspects of their performance. Individuals are encouraged to self-evaluate because self-assessment is an effective tool for changing their behavior and becoming more independent (Hargreaves, 1995; Osterman & Kottkamp, 1993; Schroeder, 1996; Simpson, 1966; Sutherland & Wehby, 2001). In education, the ability to correctly self-evaluate enables teachers to find strengths and weaknesses in their teaching and modify practice when needed (Ovando, 2001; Wagner, 2006). By identifying these areas, teachers can work to make improvements and change their behavior (Keller, Brady, & Taylor, 2005).

Self-assessment has been widely recognized as extremely efficient, particularly in the field of education (Black & William, 1998; Irvine, 1983; Kostka, 1997; Yarbrough, Wapnick, & Kelly, 1979). Through self-assessment, teaching weaknesses can be quickly eliminated and behaviors that are desired can improve and occur on a more frequent basis (Cheney, 2005; Simpson, 1966; Sutherland & Wehby, 2001). As Attinello and Lare (2006) state:

> teachers should be expected and encouraged to engage in similar activities to regularly reflect on their teaching practices, to identify their strengths and weaknesses, and to determine areas of needed improvement - improvement that will have a direct impact and significant influence on student learning. (p. 47)

If self-assessment is to be a successful force in the education profession, it must be an on-going process (Chaves, Baker, & Chaves, 2006; Nitz, 1982). As teachers self-evaluate, they learn to view themselves with a critical eye. The more they participate in

the self-reflective process, the easier it becomes to identify behavior discrepancies and resolve them.

Self View

Results from self-assessment studies seemed to be influenced by how individuals perceive themselves, their work ethic, and experience in their field (Abbott, 1965; Boud & Falchikov, 1989; Combs, 1964; Shepard, 1979; Smith & Kight, 1959; Topping, 1998). These studies found that individuals either perceive themselves in a healthy positive manner or in a negative manner. This self-view or self-acceptance is the fulfillment that the individual achieves, and is often associated with a healthy lifestyle (Shepard, 1979). A study by Fuller and Manning (1973) found that the individual most likely to benefit from self-assessment is one who has high self-esteem and a good image of themselves. The way an individual views himself or herself can have significant impact on the results of self-assessment. If each individual sees himself or herself differently, then accommodations must be made to account for these differences.

Self-improvement

A major component of the self-assessment process is the assumption that self-evaluators are striving to self-improve (Carr, 1977). They are psychologically healthy and seeking self-fulfillment through the evaluation process (Perlberg, 1983). Recognizing that behavioral changes come from the inside out, they are eager to identify any discrepancies and work to modify the behavior and eliminate the discrepancies (Bergee & Cecconi-Roberts, 2002; Cheney, 2005; Fuller & Manning, 1973; Lethco, 1999; Perlberg, 1983).

A significant portion of the literature deals with the effects of self-assessment on teaching performance (Bergee, 1993; Fuller & Manning, 1973; Keller et al., 2005;

Simpson, 1964; Yarbrough et al., 1979). At both the public school level and in higher education, self-assessment continues to be an important component in increasing the effectiveness of performance. The reflective process allows teachers to continually set and work to attain goals that over time change their performance (Donnelly, 2005; Wagner, 2006). When they are able to see themselves in a reflective manner, they become more likely to gain a realistic self-view of their teaching (Smith & Kight, 1959). Once this has been attained, performance can easily be positively changed.

In studies by Keller et al. (2005) and Sutherland and Wehby (2001) it was found that the use of self-assessment allowed teachers to increase their use of praise statements. Simpson (1964) found that self-assessment significantly improved collegiate teaching. In the music classroom, Worthy (2005) believes that self-assessment with ample practice and strict guidance can help teachers to attain more effective and efficient rehearsal techniques. For these changes in performance to occur, reflection or self-assessment should occur as often as possible (Stegman, 2007). The more time spent in these reflective actions, the more accurate the self-realization will be (Fuller & Manning, 1973).

Although a significant portion of the literature has found benefits in self-assessing, some research cautions its use. Colwell (1995) found that self-assessment failed to change intensity behaviors in teaching episodes while Chaves et al. (2006) found that self-assessment is a very private activity and can have problematic effects on students if they are overly critical of themselves. Because of varying results it is recommended that further research explore self-assessment and its effects on the individual.

The Process

A key component of self-assessment is the ability of an individual to know how to accurately self-reflect or undertake the process of self-assessment (Airasian, 1995; Simpson, 1964). Self-assessors are making judgments by applying different standards and criteria (Barnes, 1998; Falchikov & Goldfinch, 2000). A recurring problem is that they often have little to no training in assessing themselves. When there is minimal training, several variables can arise that can skew the results. These variables can include differences in the individual personality, the teaching and evaluation setting, or in the perceived expectations for completing the self-assessment. Carr (1977) found that students having no guidelines often assessed themselves much higher or lower than their actual performance.

The first step in understanding how to correctly self-evaluate is grasping the evaluation framework. This includes an understanding of methods of inquiry for personal behavior, an ability to identify both products and processes, and an ability to maintain focus and follow a structure (Bergee, 1993; Kostka, 1997; Stout, 1989). One of the main recommendations is the development of a behavioral checklist. Although difficult to develop, this checklist can provide specific guidelines in the evaluative process and keep a specified focus for those completing the checklist (Bergee, 1993; Carr, 1977; Colwell, 1995). Once the behavior checklist is developed, individuals must be trained in accurately completing it. This includes training in correctly identifying and recording behaviors and in comparing what they observe to a specified standard (Barrett, 1983; Boud & Falchikov, 1989; DiGangi, Maag, & Rutherford, 1991; Duke & Prickett, 1987; Irwin & Bushnell, 1980).

Wagner (2006) believes that the most valid results come when an individual collects data from a number of sources. Once completed, the reflective practice becomes data driven and multiple means of collecting the data can be used to check for reliability and validity of the data. These can include oral, written, verbal, and nonverbal methods (Perlberg, 1983). In higher education, it is up to the institution's teacher training program to provide the necessary instruction in these self-assessment methods.

Teacher Training

Self-assessment in education can take on a variety of forms: students examining their work, teachers modifying their own teaching performance, and preservice teachers using it as a teacher training tool. Many institutions include self-assessment as an important, though not mandated, component of their training sequence. In Canada, for instance, all teacher training institutes are expected to aid prospective teachers in developing self-analysis skills (Carr, 1977).

Developing as a teacher is highly dependent on the ability to analyze one's teaching practices (Stegman, 2007). Teachers experienced in the art of self-assessment have reported that new life is always being breathed into their teaching. Through the self-assessment process these teachers are able to describe in detail their lesson plans and intentions, explain reasons behind their actions and decisions, and adjust their practices when necessary (Clarke, 1958; Sawyer, 2001). All of these things are made possible by accurate training in the art of self-evaluating.

Institutions using self-assessment techniques in their teacher training programs have found many positive outcomes. These include: tighter connections between theory and practice, improved instruction, and greater self-realization (Claudson, 1969;

Hourigan, 2006; Keller et al., 2005). Ideas for connecting theory and practice as noted by Hourigan (2006) include:

> observing a peer or an instructor, either live or by videotape, and critiquing the student's own effectiveness through evaluation and discussion - these observations and reflections appear to enhance the ability of undergraduates to evaluate themselves as music teachers. (p. 36)

Other ways of connecting theory with practice are completing self-assessments with a coaching instructor or supervisor (Worthy, 2005). It must be noted that caution must be taken that the supervisor functions in a non-threatening capacity, encouraging dialogue instead of monologue. Attinello and Lare (2006) remark that this will allow the teacher to spend a greater amount of time in discussion and reflective thought, thus allowing him or her to improve and grow at an expedited rate.

Other methods recommended in training teachers to accurately self-assess include students developing their own case studies using collected data to make teaching decisions (Hourigan, 2006; Keller et al., 2005). Case studies allow them to analyze the teaching of others either realistically or hypothetically, which in turn helps them apply what they learn to their own self-reflection skills (Sterling, 2008). Making teaching decisions by collecting data from a variety of sources makes it possible to gather more reliable and valid data than is available from just a single source.

The development of self-assessment techniques seems to be an important component in the training of future educators. It must be remembered that a strong conceptual base for evaluating should be established early in the training process (Fuller & Manning, 1973). Students must be able to assess their progress, learn to identify

behaviors, make behavioral changes as needed, and work to take responsibility for their own learning (Boud & Falchikov, 1989; Clarke, 1958; Gunter & Reed, 1996; Perlberg, 1983; Schulz, 2005). The teacher training institution should recognize that there are multiple methods for collecting self-assessment data and that students must be taught to accurately view themselves while seeing their skills and performance objectively (Kostka, 1997).

Compared to Other Observers

Self-assessment, when compared to other forms of evaluation, often yields differing results (Bergee, 1993; Boud & Falchikov, 1989; Bullough, 1989; Carr, 1977; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Colwell, 1993; Duke & Prickett, 1987; Fuller & Manning, 1973; Keller et al., 2005; Kerrins & Cushing, 2000; Kuhn, 1991; Madsen, 2003; Stegman, 2007). Research has revealed significant differences in the evaluation of self- and instructor assessments, self- and outside assessments, the assessment of novice and experienced teachers, and the assessment of strong and weak students. Typical findings were that observers are selective in their perception; the better students tend to underrate themselves and are more critical of themselves. Students who are weaker often overrate themselves and are much less critical in their self-assessments. How can these differences be accounted for, and what methods or combination of methods can be utilized to minimize these discrepancies?

Professional Development

One goal of training preservice teachers to accurately self-reflect is to instill in them the habits that will be needed when they become practicing teachers. Practicing teachers are expected to continue learning through professional development

opportunities. Learning how to accurately self-assess can be a very effective means of professional development (Manfredo, 2006). Wagner (2006) suggests that through the actual process of self-assessment, a well-defined professional development plan can be created. When a plan is created and goals are set, the individual is able to take ownership of their teaching and continually develop as a professional. This in turn, allows the educational system as a whole to progress (Donnelly, 2005). Through these methods teacher evaluation becomes a diagnostic tool and promotes pedagogical reform (Chen, Burry-Stock, & Rovengno, 2000). If the educational system is to reach this point, self-assessment training must begin with preservice teachers.

Self-Assessment Methods

Self-assessment is an active learning process. Active learning implies that the participating learner is "doing". It is well documented that individuals who are directly involved in the learning process learn more quickly and retain knowledge for a much longer time (Astin, 1975, 1984; Fletcher, 2005; Natriello, 1984). This process of "active learning stimulates higher-order cognitive processes such as critical thinking and analysis" skills, thus creating lasting bonds in the brain (Phillips, 2005, p. 82). As an active learning process, self-assessment continually uses these higher-order processes and encourages their development and improvement (Sterling, 2008).

To this point self-assessment methods have been only briefly mentioned. A deeper examination of these methods is warranted. The four identified methods are reflective dialogue, self-reports, portfolios, and journal writing. The first method of reflective dialogue is the process of "talking out" an evaluation moment. The self-assessing teacher may discuss the evaluation with an administrator or cooperating

teacher. Stegman (2007) found that the use of a cooperating teacher makes it more likely that the self-assessing teacher will critically reflect on his or her teaching. The following five elements were found to enhance the reflectivity of student teachers: regular dialogue meetings, allowing the student teacher to initiate the course of the dialogue, modifying questions to meet the student teacher's individual needs, the issuing of further questions by the cooperating teacher, and the cooperating teacher guiding the student to deeper levels of reflection (Stegman, 2007). This view is much different than the approach taken in the 1980s and 1990s where evaluation was driven by administrative rather than individual teacher needs (Ory, 2000).

The second method is self-report. This method allows individuals to report, in a reflective manner, on their teaching. Usually the self-report comes in the form of a written or typed paper listing observations made by the individual. Sometimes a prescribed form is utilized, with the self-reporter answering the questions and often ranking aspects of his or her teaching using a Likert-type scale. This method of self-assessment is often the least expensive and quickest of the other methods. Turrentine (2001), however, found that the introduction of complex behaviors may compromise the accuracy of the evaluation.

The third method is the use of portfolios, collections of many evaluations and other artifacts collected over an extended period of time. The portfolio has been found to promote the self-assessment process and development of teaching skills (Donnelly, 2005). It must be noted, however, that portfolios can be difficult to keep track of and evaluate (K. D. Peterson, Stevens, & Mack, 2001).

The fourth method, journal writing, involves the keeping of a detailed log of teaching reflections. Over time, the process of writing out one's own reflections and thoughts allows teachers to identify needed changes in their teaching behaviors and gives them the ability to make those changes. Journal writing is an effective means of self-assessing, learning, and increasing the participatory role of the student in the learning process (C. M. Clark, 1988; Fendler, 2003; Liu, Lin, & Yuan, 2002). Often individuals write about their ideas, feelings, and changing perspectives. It "is a popular method of promoting exploration and facilitating reflection on learning and new experiences within the context in which the learning unfolds" (Gillis, 2001, p. 49). Hourigan (2006) found that self-assessment through journal writing could enhance the early teaching experience.

In conclusion, the process of self-assessment can be an important component in the training of preservice teachers and as professional development for practicing teachers. Although some research recommends caution in the use of self-assessment, the majority of studies promote its benefits.

Videotape

One method for aiding teachers in the process of self-assessment is the use of videotape. Highly researched in the 1960s and 1970s when the technology was new; videotape has shown conflicting results (Albert & Hipp, 1976; Bedics & Webb, 1971; Carr, 1977; DeBacy, 1969; Fuller & Manning, 1973; Yarbrough et al., 1979). Since then, isolated studies have continued to explore the use of videotape in teacher evaluation (Bergee & Cecconi-Roberts, 2002; Perlberg, 1983; Turrentine, 2001).

Benefits - General

There are many benefits to recording oneself using video equipment (Frager, 1985; Pailliotet, 1995; Storeygard, 1995). Video playback allows evaluation and learning to be shifted from a mostly supervisor-led activity to one that is more individual-led (Fuller & Manning, 1973). Videotaping allows teachers to continually replay and re-analyze their performance (Burrack, 2001; Perlberg, 1983). This helps them to identify exact discrepancies between what was intended and what actually took place, as well as to better understand their belief systems and how those beliefs influence their teaching aims and goals.

One of the most credible validations for the use of videotaping is its ability to provide feedback (Bloom, 1969; Carr, 1977; Olivero, 1966; Perlberg, 1983; Winn, 1974). Videotape feedback can provide four things. First, it offers a specific framework for self-reflection. The individual can isolate him or herself and focus solely on the review of the video without other distractions. Second, it can validate other evaluation sources and allow comparisons between self-assessments and those of peers and supervisors. Third, it can be used to supplement other strategies. Fourth, it can motivate change. If teachers who may discount comments made by other evaluators are given an opportunity to view their own teaching on video, they may be more likely to implement the necessary changes.

Bergee and Cecconi-Roberts (2002) found that videotaping enhanced the ability of music students to evaluate themselves. Analyzing video can change all kinds of behaviors, including the more complex ones. By enabling individuals to see themselves as others see them, it provides a greater opportunity to learn (Albert & Hipp, 1976; Penny

& Coe, 2004). Beck, King, and Marshall (2002) found that using videotape allowed individuals to develop their own interpretation of events. Once the events are identified, they are able to change their behavior and walk away with new knowledge (Marion & Samaniego, 1981; Perlberg, 1983). Kpanja (2001) found that these groups and others using video made significant improvements over groups not using video.

Benefits – Teachers

Videotaping for self-assessment is especially useful for teachers. Research by Albert and Hipp (1976), Cassidy (2005), Kpanja (2001), Marion and Samaniego (1981), and Penny and Coe (2004) have found that videotaping can be used: to improve teaching performance, as a quick method of spotting discrepancies, and combined with other methods to produce an improved teaching product. Teaching performance attributes such as verbal delivery of material, maintaining student interest, reinforcing subject matter, and drilling techniques have improved (Gonzo & Forsythe, 1976; Stoll, 1973). To spot teaching discrepancies, individuals are often encouraged to view the videotape multiple times. By viewing the videotape repeatedly, teachers learn to identify and analyze their teaching episode and thus discover any errors in their teaching. (Bedics & Webb, 1971; DeBacy, 1969; Manfredo, 2006).

In music research similar results have been reported (Barnes, 1998; Gonzo & Forsythe, 1976; Price, 1992; Stuart, 1979; Taebel, 1980; Tjornehoj, 2001; Worthy, 2005; Yarbrough et al., 1979). Studies have noted improvement in teaching by music therapists and music teachers in many different areas. These include one's ability to better identify performance errors, time allocation skills among teachers, and effectiveness of beginning conductors.

Teacher Training

With the benefits identified, it is important to examine how teacher training programs can use the videotape in their preservice programs. Perlberg (1983) notes that video can be used in teacher training in a multitude of ways. These include: the taping and review of practicing teachers as models to the preservice teacher, the taping of the preservice teacher for self-assessment, the taping of the preservice teacher for supervisor and/or peer critique, and helping other training techniques become more effective (Kpanja, 2001; Worthy, 2005).

When preservice teachers examine a videotape they are able to more accurately reconstruct what happened during the teaching setting than if they were trying to reconstruct without the videotape (Albert & Hipp, 1976). With the capabilities of multiple playback, the video serves as an effective tool for isolating and more easily identifying teaching mistakes and discrepancies (Cassidy, 1990; Pailliotet, 1995; Sherin, 2000). It allows preservice teachers to more easily recognize behaviors that occur during the teaching segment (Penny & Coe, 2004). The preservice teacher using videotape soon becomes more independent and relies less on feedback from external sources (Price, 1992). Self-assessments of those using videotape usually come more into line with the marks of other observers. Although some researchers disagree, Bloom (1969) notes that seeing oneself on video does not seem to bother the observing participant.

Several studies have found positive correlations between the use of video and specific teaching behaviors (Fuller & Manning, 1973; Price, 1992; Thompson, 2007). These studies found increased ability of participants to understand their prior beliefs, identify preconceived notions about their teaching, give quality feedback to students, and

give less threatening feedback to other classmates. It must be noted that the video should be used to focus only on those behaviors that the teacher has the ability to control (Penny & Coe, 2004).

The use of videotape in teacher training is useful to not only the preservice teacher but to professors or supervisors as well. Many collegiate programs, especially those training large numbers of preservice teachers, often are constrained by time factors. The use of videotape can help reduce evaluation time by allowing the review of the video and discussion to be held in one location (Marion & Samaniego, 1981).

The repeated viewing of the videotape allows individuals to process what they see in a variety of ways (Beck et al., 2002), although Holzman (1969) found that the greatest change and impact came from the first viewing of the taped episode. Other research techniques of playback include the use of longer episodes rather than shorter ones and providing the preservice teacher with a specific viewing list or criteria; playback should occur as soon as possible following the teaching episode (Beck et al., 2002; Benne, Bradford, & Lippitt, 1964; Perlberg, 1983).

Problems with Videotape

Although the majority of the research supports the use of videotaping, the literature contains some cautions. Kpanja (2001) believes that the use of a videotape can often lead the user to act in an artificial manner, partly because knowledge that one's performance is being taped can cause stress and anxiety. Other problems with videotaping include the length of time it takes to view and review one's teaching, self-image issues that result from seeing oneself on film, technical problems that arise when the equipment malfunctions, and a lack of impact on learning (R. E. Clark, 1983;

Colwell, 1995; Eckrich & Widule, 1994; Logue, Zenner, & Gohman, 1968; Perlberg et al., 1971; Steward & Steward, 1970). Since video recording has the potential for harm, it is not recommended for group playback. Fuller and Manning (1973) recommend viewing alone or with a supervisor or teaching partner.

Besides personal confrontational issues, some studies have found that the use of videotape has little or no effect. Penny and Coe (2004) found that little benefit resulted from pairing observation with videotaping. Kpanja (2001) and Perlberg (1983) noted that effects, particularly individual behaviors, remained unchanged whether videotaping was used or not. Gonzo and Forsythe (1976) found that videotaped excerpts of student and teacher interactions were invaluable in the training of music teachers. Further research is needed to clarify these conflicting research results.

Peer Critique and Assessment

Peer critique and assessment can be another beneficial technique for use in the teacher training program (Barber & Lewis, 1986; Garmston, 1987; Gordon, 2002). In recent years peer critique and evaluation have grown in popularity (Bernstein, Jonson, & Smith, 2000). Peer critiques are usually completed by a group (Falchikov, 1999; Zepeda, 2007). Peer critique is highly reliable and practical in the higher education setting (Anderson & Freiberg, 1995; Darling-Hammond, 1988; Lieberman, 1988; Liu et al., 2002). Gillis (2001) notes that peers in similar situations are often helpful to each other since their teaching context is usually similar. Unfortunately teachers often have little opportunity to discuss and critically reflect on their teaching practices (Goldstein & Noguera, 2006).

Benefits and Outcomes

The use of peer critique and assessment offers many benefits. These can occur at both the individual and the group level. At the individual level, peer critique has been found to increase students' ability to self-assess and critique their teaching practices, develop critical thinking skills, produce better reflection through dialogue with others, strengthen leadership competencies, and increase the ability of the individual to identify leadership behaviors in themselves as well as others (Bergee & Cecconi-Roberts, 2002; Lepard, 2002; Schulz, 2005; Stegman, 2007; Topping, 1998; Turrentine, 2001).

Peer critique has been found to promote group learning and to function as a tool for decreasing discrepancies between supervisor and peer marks (Falchikov & Goldfinch, 2000; K. D. Peterson, Kelly, & Caskey, 2002; Ward, 2005). At the public school level, peer critique has become an important component in evaluating teacher effectiveness (Bergee, 1993). It provides more frequent opportunities for teachers to self-reflect. As a result of this constant interaction and critique, teachers are less severe with each other and principals need to spend less time in the observation task (Goldstein & Noguera, 2006; Hirschfeld, 1968; Schulz, 2005).

Peer assessment "involves students directly in the learning process and may promote a sense of ownership, personal responsibility, and motivation" (Topping, 1998, p. 256). Often students are better able to identify aspects of their own teaching when they have had frequent opportunities to critique others. Comparing one's own technique and abilities to others is beneficial, and satisfaction comes from working with others and being open to their ideas (Cheney, 2005; Sterling, 2008).

Uses of Peer Critique

Quality teaching in today's society requires teachers to be active learners. Peer critique has been found to be a useful tool in promoting this type of learning (Chaves et al., 2006; Topping, 1998). Use of this model allows teachers to hold each other accountable (Goldstein & Noguera, 2006). Working with other teachers allows them to discuss ways in which theory and practice can be combined (Hourigan, 2006). After working together teachers are usually better able to assess themselves. Turrentine (2001) found that 83% of the time, self-assessments closely resembled peer critiques.

Peer critique can also be an important part of teacher education. It can give students a better understanding of the assessment process and allow them to become more competent teachers (Bergee, 1993; Chaves et al., 2006; Keig, 2000; Topping, 1998). Peer critique allows for greater and swifter feedback. It "provides a setting for teachers to discuss problems, share ideas, help prepare lessons, exchange tips, examine student work, and provide support to one another" (Berube & Dexter, 2006, p. 12). It also:

> supports small groups of teachers working together to grow professionally, and self-directed development – involves teachers setting goals, obtaining feedback from peers and experts, and assessing their own progress. (Berube & Dexter, 2006, p. 13)

Consultation—that is, group discussion time—is considered an important part of the peer critique process. It is a good method for cooperative development (Berube & Dexter, 2006). Very often it is used for teachers looking to improve specific skills or their

teaching in general (Bergee, 1993; Penny & Coe, 2004). The process of consultation is usually a non-threatening collaborative approach that is highly structured.

How to Use Peer Critique

An important part of the peer critique process is setting parameters and correctly structuring the process. Falchikov and Goldfinch (2000) recommend that a small number of peers be placed in each group. They also state that peer critiques are likely to be more valid in higher-level classes. Bandura (1977) recommends that group members be close in age, personality, and achievement.

A second important part of the peer critique process is that specific, detailed non-threatening feedback should be continually provided (Bergee, 1993; Falchikov & Goldfinch, 2000; Sawyer, 2001). To obtain this, it is best advised that completion of evaluations be kept anonymous (Ward, 2005). Martin (2004) recommends that this method is best used in the early program years. Typically specific feedback is provided by peers who have decided in advance what the focus of the observation should be (Schulz, 2005). It is shared evaluation (Donnelly, 2005). Peers are able to decide the focus of the observation by predetermining the judging criteria, how they will assess the teaching episode, and how it will be graded (Liu et al., 2002). If these parameters are in place, the peer critiques and assessments are more likely to resemble that of the supervisor (Chaves et al., 2006). This will in turn create a more unified view for the individual receiving the feedback.

Peer critique can be an effective tool for changing an individual's behavior. When groups of peers are provided with expectations and specific evaluation and feedback procedures, the critiques are likely to be less conflicting and provide greater opportunity

for this behavioral change. The supervisor or an assigned coach can be used to provide these procedures as well as help to the participant when needed (Goldstein & Noguera, 2006).

Problems with Peer Critique

Peer critique is not without problems. Topping (1998) has identified three. The first is that weaker teachers often see peer feedback as being inaccurate. The second is that studies on the use of peer critique have found little improvement in the presentation and confidence skills of individual participants. The third is that self-assessments were often found to be more reliable than those of peers, possibly because peers often use more positive descriptors than does the individual (Chaves et al., 2006). As with the video research, conflicting results warrant further study.

Cooperating Teacher and Supervisor

Traditionally, the cooperating teacher or supervisor has functioned as the guiding force behind behavioral awareness and teacher improvement (Stodolsky, 1984). In this role, the supervisor is expected to function in a non-threatening manner. When this occurs, the individual is more likely to accept the feedback, whether negative or positive, and apply it to their teaching (Schmidt, 1994). As Berube and Dexter (2006) state: "supervision is not about judging a teacher; it is an opportunity to facilitate dialogue around what is working and what is not working in the classroom" (p. 16). Simpson (1966) recommends accepting criticism as valuable and helpful information rather than as a personal attack. The more comfortable the teacher feels when speaking with an administrator, the more likely he or she is to be open to suggestions and needed changes.

During the dialogue following an observation, the supervisor functions as a guide. He or she helps students identify problems with their teaching, appropriately express their emotions, and connect theory with practice (Gillis, 2001; Merkley & Hoy, 1985). All dialogue led by the supervisor should work toward the common goal of developing the needed skills for the teacher to successfully function in the current teaching environment (Clarke, 1958). For this to occur, adequate supervision must be continually provided to the teacher.

Several cooperating teacher and supervisor studies have made recommendations for the evaluation process (Bailey, 1981; Berube & Dexter, 2006; Hollingsworth, 1989; Marion & Samaniego, 1981; Marshall, 2005, 2008; Panhorst, 1971; Schueler & Gold, 1964). These include: having two rather than one supervisor present at the post-observation conference, creating an observation plan with the teacher that identifies specific goals and focuses for the observations, giving feedback promptly following an observation, using videotape to refine observation skills, and providing a written detailed summary of observation notes.

Although the majority of cooperating teacher and supervisor research is positive, some is negative (Kerrins & Cushing, 2000; Pajares, 1992; Stodolsky, 1984; Stout, 1989). Studies have found that supervisor dialogue and conferences often have little impact on the teacher's performance; changing the teacher's beliefs and commitments can be damaging; expert, novice, and supervisor markings differ; and providing only surface level feedback often does not allow teachers to form a framework that they can later apply to other situations. Further research by Kottkam, Provenzo, and Conn (1986)

and Shedd (1985) found that the individual being evaluated often dismisses the supervisor's feedback as not useful.

"The role of teachers and faculty in supporting and guiding students to become reflective practitioners cannot be overemphasized" (Gillis, 2001, p. 58). The evaluation process must remain non-threatening while challenging the teacher to recognize and change discrepancies in their teaching. The goal of all supervisor evaluation should be to develop teachers who can evaluate their own teaching (Verrastro, 1975). Teachers who learn to accurately self-assess can more easily recognize discrepancies in their teaching and work to correct them.

General Assessment

Assessment of teachers often involves the capture of only a small portion of a lesson, giving an incomplete picture of the teacher's abilities (Marshall, 2005). Compounding this problem are the different practices found at each institution (Blanton, Sindelar, & Correa, 2006). For the preservice teacher, assessment is even more problematic. Preservice teachers are often evaluated in the practicum setting on techniques that they have just learned. These evaluations usually focus on a technical aspect of their teaching, even though this accounts for only a small portion of their teaching knowledge (Schulz, 2005). These micro-evaluations often carry little weight (Marshall, 2005). Because of these issues and problems, the process of evaluation is often perceived as having little impact on instruction and improvement of practice (K. D. Peterson, 2000).

If the assessment is to be credible, valid, and helpful to the practicing teacher, what should be assessed? Falchikov and Goldfinch (2000) recommend that evaluations

focus on relatively few dimensions rather than trying to assess a multitude of dimensions. In order for these dimensions to be thoroughly understood, preservice teachers must have the opportunity to encounter, early on, the conceptual framework for teaching and learning. This will help them in their observation and evaluation process (Beck et al., 2002). Observations can examine a number of things including: nonverbal behavior, lesson content and sequence, attention to classroom behavior issues, wait time between questions, supportive statements, and/or verbal behavior, such as delivery skills (Hamann, Baker, McAllister, & Bauer, 2000; Stroh, 1968).

If teaching behavior is expected to change, quality assessment and evaluation must have good feedback that comes from a highly specific uniform analysis of the evaluation data (Bergee, 1993; "Evaluating Teaching," 1963; Geertsma & Reivich, 1965; Karasar, 1970; Morse, Kysilka, & Davis, 1970; Pease, 1972; Schulz, 2005; Yarbrough, 1987). Evaluation data can come from sources such as individuals completing direct observation or alternative forms such as tests, self-critiques, and portfolio creations (Blanton et al., 2006; Guthrie, 2005).

The next question that arises from general assessment is: Who should provide the feedback? Should one person or a multitude of individuals be utilized? Should assessment and evaluation be limited to supervisors or should the individual and peers be involved as well? Several researchers recommend that multiple evaluators be used (Perlberg, 1983; K. Peterson, 2004; Phillips, 2005; Sterling, 2008). Utilizing multiple sources often allows greater individual acceptance of the data, more opportunities to improve teaching, and a more uniform view of evaluation data. The use of multiple sources is not without skeptics. One researcher found that it is very difficult to compile

uniform data from a multitude of sources ("Evaluating Teaching," 1963). Perlberg (1983) remarks that correlating the sources is essential. One suggested way in which correlation can be more easily obtained is the use of a Likert-type scale on the evaluation form (Boud & Falchikov, 1989; Marshall, 2005).

Many variables come into play when assessing the preservice teacher. It is imperative that the strategies and framework for assessment be thoroughly examined and that only the most effective techniques and strategies be utilized in these evaluations.

Teacher Education

Collegiate teacher training programs have a very significant impact on preservice teachers. However, current research on teaching training has revealed several problems (Schulz, 2005; Stout, 1989). First, current methods of teacher training practice do not allow preservice teachers to experience the full range of practicing teacher duties. Second, students are rarely given the opportunity to self-reflect and learn self-directing/learning skills. Third, teacher certification rarely requires evidence of teaching performance (Airasian, 1995). These problems are generally attributed to a misaligned teacher education framework and a lack of uniform evaluation and assessment procedures. Clarke (1958) found that most institutions were evaluating based on: participation of the student in conferences, logs and diaries completed by the student, and students' individual lesson plans. Wilderson, Manatt, and Maughan (2000) state:

> traditional teacher evaluation has focused on a process whereby a single administrator conducted a limited number of classroom observations, basing recommended improvements on these limited contacts. (p. 179)

A great deal of research has presented recommendations for the restructuring and revision of current teacher education practices (Fuller & Manning, 1973; Hamann, Lineburgh, & Paul, 1998; Panhorst, 1971; Schulz, 2005). These recommendations include the use of video- and audiotape, supervisor feedback, peer collaboration and greater attention to discussion, new courses on teacher effectiveness, peer teaching, supervisor reflection time, class discussions, and alternative evaluation materials such as self-reports or portfolios (Keller et al., 2005; Kretchmer, 2002; Lunenberg & Willemse, 2006; Stegman, 2007; Stout, 1989; Worthy, 2005).

Other researchers recommend that changes in the teacher preparation process should be ongoing. Methods should be created, tested, and reevaluated (C. M. Clark, 1988). Preservice teachers bring certain beliefs with them that drive their views and decisions. Each individual will be different (Burrack, 2001; Butler, 1999). Since society and thus the student population are continually changing, it makes sense to look at teacher education and evaluation as a continuous progression (Tyler, 1949). Continuous change in the student population means continuous change in the practicing teacher population as well. Cassidy (1990) states:

> if teacher training is to have an impact on the classroom skills of novice teachers, attributes of successful teachers must first be identified and then defined in such a manner that they can be taught and evaluated. (p. 164)

These traits can only be identified through study and evaluation of the current teaching population. Once identified, they should be used to modify existing teacher training programs. These modifications should allow the program to ensure that students are

gaining the appropriate knowledge and skills to be able to self-determine the "how- to's" of teaching (Walls, Nardi, & Von Minden, 2002).

In music education, research has also focused on teacher education, training, and evaluation (Duke & Blackman, 1991; Maranzano, 2000). Hamann, Baker, McAllister, & Bauer (2000) found that both practicing teachers and preservice teachers place greater emphasis on their teaching skills than their musical skills. They acknowledge that music teachers must be trained to present material in an effective manner. Effective presentation can involve: the preservice teacher presenting a task, having the student interact with the specific task, and providing relative feedback (Price, 1992). Thompson (2007) remarks that many of the teacher training programs in music lack cohesiveness. This lack of connection makes it hard to have wide effect in music education. Teacher training programs must prepare individuals to be competent and effective teachers (Lethco, 1999). Music education as well as general education must work to create quality methods for teaching students. This can only be accomplished through evaluating, testing, and retesting current trends and methods of training and teacher assessment.

Implications for This Study

In summary, a review of the related literature reveals a plethora of research in the field of evaluation and assessment. Evaluation by peers and supervisors as well as self-reflection through the use of videotape have been explored. Although there are a large number of studies on evaluation and assessment, results are varied. In an age of accountability, with demands on teachers

continuing to increase, it is important to test and retest evaluation and assessment strategies to determine which technique or combination of techniques is most effective for increasing teacher effectiveness and meeting the new levels of accountability.

CHAPTER 3

METHODOLOGY

Since the late 1800s, when teacher evaluation and assessment began, researchers have continued to explore a variety of evaluative methods and techniques that can produce the most effective teachers. It is generally agreed that self-assessment, peer critique, and supervisor evaluation are the three most important modes for aiding in teacher development and improvement. Unfortunately, research has revealed conflicting data when self-assessment, peer critique, and supervisor evaluation have been tested. Thus, the purpose of this study was to determine the effect of videotape analysis and peer critique on developing the skill of self-assessment in pre-service teachers. It is hoped that isolating these specific components of evaluation will lead to a better understanding of their impact on the teacher's ability to self-assess and thus result in a clearer understanding of the benefits of evaluation in education.

Participants

Participants (N=40) were selected from two undergraduate music education courses in the School of Music at a large southeastern public university. The first class consisted of participants working to complete their degrees in music education. The second class consisted of participants working to complete education degrees in a variety of fields (e.g., special education, mathematics, middle school). The majority of the participants were in their junior year of college and were to begin student teaching within

one calendar year. Any gender, age, and ethnicity differences resulted from random selection procedures.

Design

Participants (N=40) were randomly divided into four groups: three experimental and one control. The Control Group (n=10) performed a self-assessment following each teaching episode (Table 3.01). Experimental Group I (n=10) self-assessed with the aid of a video recording of their lesson (Table 3.02). Experimental Group II (n=10) self-assessed with the aid of peer critiques and evaluations (Table 3.03). Experimental Group III (n=10) self-assessed with the aid of both videotape and peer critiques (Table 3.04).

Table 3.01

Control Group Design

1. Teach Lesson 1	5. Teach Lesson 3
2. Self-Evaluate	*6. Self-Evaluate*
3. Teach Lesson 2	7. Teach Lesson 4
4. Self-Evaluate	*8. Self-Evaluate*

Table 3.02

Experimental Group I Design

1. Teach Lesson 1	7. Teach Lesson 3
2. Self-Evaluate	*8. Self-Evaluate*
3. Videotape Analysis	9. Videotape Analysis
4. Teach Lesson 2	10. Teach Lesson 4
5. Self-Evaluate	*11. Self-Evaluate*
6. Videotape Analysis	

Table 3.03

Experimental Group II Design

1. Teach Lesson 1	7. Teach Lesson 3
2. Self-Evaluate	*8. Self-Evaluate*
3. Peer Critique Review	9. Peer Critique Review
4. Teach Lesson 2	10. Teach Lesson 4
5. Self-Evaluate	*11. Self-Evaluate*
6. Peer Critique Review	

Table 3.04

Experimental Group III Design

1. Teach Lesson 1	7. Teach Lesson 3
2. Self-Evaluate	*8. Self-Evaluate*
3. Videotape Analysis & Peer Critique	9. Videotape Analysis & Peer Critique
4. Teach Lesson 2	10. Teach Lesson 4
5. Self-Evaluate	*11. Self-Evaluate*
6. *Videotape Analysis & Peer Critique*	

All four groups began the study with a personality profile. The *Internet Personality Inventory Survey* (Appendix B) is an online personality profile that measures an individual's level in each of the five personality domains: extraversion, agreeableness, conscientiousness, emotional stability, and openness. The test is designed to report normal differences in personality. It consists of sixty short questions with six possible Likert-type scale answers ranging from very inaccurate to very accurate. Participant averages, including high and low scores for each of the five domains, can be found in Figure 3.01.

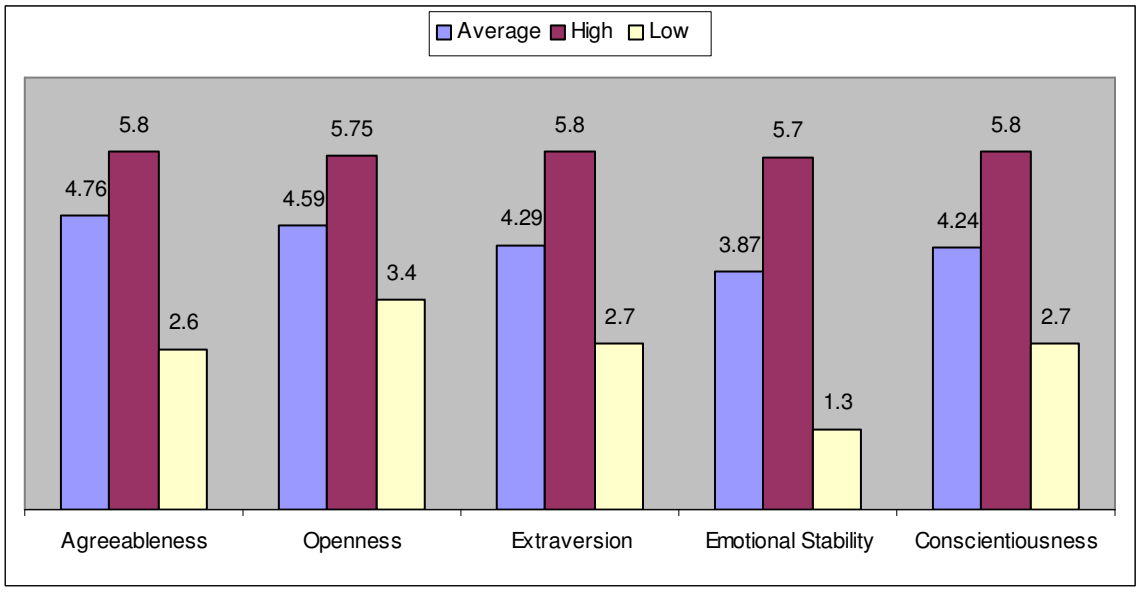


Figure 3.01. Averages, High and Low Scores for Personality Profile Domains

Following the personality profile, participants completed the pretest. The pretest was identical for all four groups, consisting of teaching a ten-minute mini-lesson. Upon completion of the teaching segment, the participants filled out the first self-assessment form (Appendix D). The self-assessment form consisted of twenty questions, short statements that assess a student's teaching. The form used a Likert-type scale. No feedback from the supervisor, videotape or peers was provided until after the participants had completed their first self-assessment.

Following the pretest and self-assessment the Control Group (n=10) received neither videotape review or peer critiques. Experimental Group I (n=10) video recorded their lesson and reviewed it following completion of their first self-assessment. Experimental Group II (n=10) examined feedback from their peers following completion of their first self-assessment. Experimental Group III (n=10) reviewed video recording of their lesson along with peer feedback after completing their first self-assessment. Peer feedback came from forms completed during the teaching of Experimental Group II and

III lessons. The forms were then gathered and redistributed to each participant in those groups.

Once the pretest was completed, each participant presented two more lessons. As with the pretest, each participant completed another self-assessment form, Appendix E for lesson two, which contained sixteen questions, and Appendix F for lesson three which contained twenty-nine questions. All four groups then taught a fourth and final lesson. Similar to the pretest, immediately following the lesson each participant self-assessed without the aid of videotape or peer critique. Evaluation form four which contained twenty-nine questions was used (Appendix G).

Pilot Study

A pilot study was administered prior to the start of the experiment in order to clarify the experimental design and to check for any uncontrolled variables. Recommendations for changes to the evaluation forms as well as questions about the study design were discussed with the pilot study participants. Appropriate modifications were made to the research design and materials.

Variables

There were several variables that had to be controlled in order to produce a credible study. The following are controls that were set in place for the experiment. Participants were selected and placed in either the Experimental or Control Group based on a table of random numbers. Each participant then completed an online personality profile (Appendix B). The appropriate observation forms were selected by professionals in the field. The four forms ranged from twenty to thirty questions that evaluated the

participants' teaching. A panel of professionals in the field examined the forms to determine their reliability and validity.

The video camera recording each lesson was set up in the same position for each participant regardless of the experimental group. A rubric was created and confirmed by three professionals in the field. The rubric was used to evaluate the responses on the prescribed evaluation forms. Peer critiques were completed during the teaching lessons. Names and any identifying means were left off. Supervisor and peer verbal comments were not allowed following the completion of each lesson.

Procedures

Participants (N=40) were randomly divided into one of four groups. Once divided, each participant completed the personality profile. The personality profile was completed online. At the conclusion of the profile, participants were given a score. They then provided the researcher with either a copy of the score or a link to the score. An example of a completed profile can be found in Appendix H. When all of the personality profiles were completed and collected, the experimental portion of the project began.

Each of the four groups taught four times. The first and the last teaching episodes were identical for all four groups. Teaching episodes two and three were different for each of the groups. The Control Group continued to self-assess without the use of videotape or peer critique. Experimental Group I self-assessed with the aid of a video recording of their teaching. Experimental Group II self-assessed with the aid of their peers' critiques. Experimental Group III self-assessed with the aid of a video recording of their lesson and responses from their peers' critiques. Upon completion of the two

teaching episodes, all the participants completed the posttest. The posttest followed the same procedures as that of the pretest. Data was then collected and analyzed.

Timetable

The experimental portion of the study began early in the spring semester. The study ran for fourteen weeks. Appendix C notes the week-by-week timetable of activities and corresponding assignments.

CHAPTER 4

RESULTS

The purpose of this study was to determine the effect of videotape analysis and peer critique on developing the skill of self-assessment in pre-service teachers. This research addressed the following questions:

1) Does the use of videotape analysis and a peer critique process result in more accurate self-assessment?
2) Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate?
3) To what extent do the self-assessments agree with those of experts?
4) Do identifiable personality traits affect accuracy in self-assessment: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience?

All of the participants (N=40) began the study with an online personality profile test (Appendix B). Each then taught four lessons with a self-assessment immediately following each lesson. Experimental Group I viewed a videotape of their lessons, Experimental Group II examined peer critiques, and Experimental Group III viewed a videotape of the their lessons and examined peer critiques.

Participant Demographics

The study began with forty-two participants. Through the course of the study a decay of two participants occurred. One participant withdrew from the class and the other decided to not be a part of the study. The remaining participants (N=40) were randomly selected from two music classes. A table of random numbers was used to place participants in one of the four groups. Participant demographics can be found in Table 4.01. All ethnic and gender differences were the result of random selection procedures.

Table 4.01

Participant Demographics

Ethnicity	*N*	*Gender*	*N*
African-American	1	Female:	21
Asian	1	Male:	19
Caucasian	38		
Hispanic	0		
Native American	0		

Question Introduction

Once all of the data was collected, data analysis began. In order to thoroughly answer each of the four questions posed in this study, both quantitative and qualitative measures were utilized. For research question (1) *Does the use of videotape analysis and a peer critique process result in more accurate self-assessment?,* research question (3) *To what extent do the self-assessments agree with those of experts?,* and research question (4) *Do identifiable personality traits affect accuracy in self-assessment: extraversion,*

emotional stability, agreeableness, conscientiousness, and openness to experience?, quantitative measures were utilized. Results for research question (2) *Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate?,* came from descriptive data collection techniques.

Research Question One

The analysis in this section provides information to answer the following research question: Does the use of videotape analysis and a peer critique process result in more accurate self-assessment?

Participants' self-assessments and the professional committees evaluations were collected once completed. All of the evaluations for each of the four taught lessons were analyzed and given a raw score derived from assigning a Likert-type scale to each of the forms questions. All of the raw scores were then converted to percentages. Once a percentage was derived, the difference between the committee's percentage and the participants' percentage was derived. This derived score represented the difference (*d*). Once the difference (*d*) was calculated, each group's average difference (*D*) was obtained. Average differences and the standard deviation (*SD*) for each group can be found in Table 4.02. Complete data for all four groups and teaching lessons can be found in Appendices I-L.

Table 4.02

Difference (*D*) between All Four Lessons for Each of the Groups

	Teaching Lesson			
	One	Two	Three	Four

Control Group				
-Average	9.186	7.275	8.398	11.477
-SD	11.250	7.220	4.969	5.602
Experimental Group I				
-Average	5.568	7.275	12.31	10.513
-SD	9.317	4.20	5.527	6.917
Experimental Group II				
-Average	10.679	7.27	11.347	11.539
-SD	10.432	6.238	5.109	7.176
Experimental Group III				
-Average	6.707	8.364	12.114	13.334
-SD	9.970	6.01	6.30	4.767

The average scores for the Control Group ($\bar{x}$ =7.275) and Experimental Group II ($\bar{x}$ =7.27) on the self-assessment were closest to that of the professionals for the second lesson. The closest score to that of the professionals for Experimental Group I ($\bar{x}$ =5.568) and Experimental Group III ($\bar{x}$ =6.707) followed the first teaching lesson. The average scores that were the furthest from those of the professionals in the field were found after teaching lesson four for the Control Group ($\bar{x}$ =11.477), Experimental Group II ($\bar{x}$ =11.539), and Experimental Group III ($\bar{x}$ =13.334). Experimental Group I's greatest difference ($\bar{x}$ =12.31) was found after teaching lesson three. Standard deviations for each of the four groups generally grew closer together after each lesson.

Teaching Lesson One

Experimental Group I had the lowest average score ($\bar{x}$ =5.568), while Experimental Group II had the highest average ($\bar{x}$ =10.679). The average scores of the Control Group and Experimental Group II were similar, with a difference of less than two points (*d*=1.493) while the average scores of Experimental Group I and Experimental Group III, were similar with a difference of slightly more than one point (*d*=1.139). Standard deviations between all four groups were similar, with a maximum difference from the highest to lowest score of slightly less than two points (*d*=1.933).

Teaching Lesson Two

The average self-assessment score differences were the most similar for teaching lesson two. The Control Group, Experimental Group I, and Experimental Group II had identical average scores ($\bar{x}$ =7.27), while Experimental Group III was slightly higher ($\bar{x}$ =8.36). The difference between Experimental Group III and the other three groups was slightly greater than one point (*d*=1.094). Standard deviations between the Control Group, Experimental Group II, and Experimental Group III were similar (*SD*=6.23-7.22) while Experimental Group II was less (*SD*=4.20). All groups were close, with a maximum difference from the highest to lowest score of slightly less than two points (*d*=1.933).

Teaching Lesson Three

All three Experimental Groups had higher average self-assessment score differences ($\bar{x}$ =11.347-12.31) than the Control Group ($\bar{x}$ =8.398) for teaching lesson three.

The difference between the Control Groups average and the highest Experimental Groups average was close to four points (*d*=3.912). Standard deviations between all four groups were similar, with a maximum difference from the highest to lowest score of nearly one and one-half points (*d*=1.331).

Teaching Lesson Four

Of the four teaching lesson self-assessments, lesson four yielded the highest average scores. The lowest average difference was found in Experimental Group I ($\bar{x}$ =10.513) while Experimental Group III had the highest average ($\bar{x}$ =13.334). All four groups had similar averages, with a difference of around three points (*d*=2.821). Standard deviations between the Control Group (*SD*=5.602) and Experimental Group III (*SD*=4.767) were similar, while Experimental Group I (*SD*=6.917) and Experimental Group II (*SD*=7.176) were similar. The maximum difference between the highest and lowest standard deviation scores of all four groups was similar, with a maximum difference from the highest to lowest score of nearly 2.5 points (*d*=2.409).

Pretest - Posttest Comparison

Final data analysis utilized a pair wise comparison analysis of variance (ANOVA) to test the change in mean scores from the pretest to the posttest for each of the four groups. Six comparisons were made: control to experimental I, control to experimental II, control to experimental III, experimental I to experimental II, experimental I to experimental III, and experimental II to experimental III. Final data analysis revealed no difference in the self-assessment markings between the four groups, F (3, 36) = 0.79, p = 0.5095. Table 4.03 presents the results of the ANOVA tests. Full data analysis can be found in Appendix Q.

Table 4.03

ANOVA on Mean Pretest-Posttest Scores Across All Groups

Source	Sum of squares	*df*	Mean sum of squares	F	Pr>F
Model	207.635	3	69.211	0.79	0.5095
Error	3169.0845	36	88.0301		
Corrected Total	3376.7195	39			

Research Question Two

The analysis in this section provides information to answer the following research question: Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate?

When the data from all four lessons was collected, all participants were asked several questions and to provide feedback. Twenty-eight (n=28) participants provided feedback while twelve preferred to not comment. Of the twenty-eight, 6 were from the Control Group, 9 were from Experimental Group I, 8 were from Experimental Group II, and 5 were from Experimental Group III. The list of questions can be found in Table 4.04.

Table 4.04

Participant Questions

1. Where you able to identify any teaching weaknesses during the course of the study?
2. If you identified teaching weaknesses, when during the study were they?

3. If you viewed a videotape of your teaching, do you feel that it enabled you to more quickly identify teaching weaknesses?
4. If you examined peer critiques of your teaching, do you feel that they enabled you to more quickly identify teaching weaknesses?

All of the respondents (n=28) said they were able to identify at least one teaching weakness before teaching their third lesson. Twenty-six of them were able to identify other teaching weaknesses following teaching lessons three and four. Of the two participants who were unable to identify any teaching weaknesses after lesson two, one was in the Control Group and the other was in Experimental Group II.

Fourteen of the respondents viewed a videotape of their teaching following each lesson. These respondents included participants in Experimental Group I and Experimental Group III. All of the respondents stated that they felt that the videotape review process enabled them to more quickly identify teaching weaknesses. One participant stated: "Videotape does not lie…it enabled me to quickly discover which nuances worked and which did not." Another respondent who used videotape remarked: "I feel like I improved much more quickly than my classmates who were not videotaped."

Respondents (n=13) utilizing peer critique presented mixed results. Nine of the thirteen stated that the peer critique process did help them identify teaching weaknesses at a quicker rate. One of the respondent's comments: "the peer critique process opened my eyes to things that I was not doing or things that I could improve upon." The other four respondents felt that the peer critique markings could not be trusted. One participant

stated: “with peer critiques it is hard to know where they [the evaluator] are coming from. They do not know your intentions so how can they provide accurate feedback?” Another stated: “there were a variety of peer marks. One peer would mark that I did something well, another would say that I did it poorly.”

Research Question Three

The analysis in this section provides information to answer the following research question. To what extent do the self-assessments agree with those of experts? In order to check for the difference between the markings of the participants’ self-assessments and that of the professionals, simple statistical principles were utilized to find the difference (*d*) for each participant (Appendices M-P) as well as the average difference (*D*) for each group.

Teaching Lesson One

Average Difference

The average self-assessment scores for teaching lesson one were similar between groups (Table 4.05). The lowest average was in Experimental Group I ($\bar{x}$ =83.409) while the highest average was Experimental Group III ($\bar{x}$ =86.817). The difference between the highest and lowest average was close to 3.5 points (*D*=3.408).

The average professional markings for teaching lesson one were slightly further apart than that of the participants’ self-assessments. The lowest average was with the Control Group ($\bar{x}$ =85.568) while the highest average was Experimental Group II ($\bar{x}$ =89.885). The difference between the highest and lowest score was *D*=4.317.

Differences between the self-assessments and the professional evaluations ranged from 0.796-6.135. Experimental Group II had the largest difference in scores (*d*=6.135) while the Control Group had the smallest difference (*d*=0.796).

Table 4.05

Lesson One Average Evaluations for Each Group

Teaching Lesson	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
One				
-Participant	84.772	83.409	83.750	86.817
-Professional	85.568	87.385	89.885	89.658
-Difference (*D*)	0.796	3.976	6.135	2.841

Individual Differences

Individual differences and how many participants fell under each point range can be found in Table 4.06. Eighteen of the total participants scored within 5 points of the markings of the professionals in the field. Four of these eighteen participants had the exact same markings as that of the professionals. Nine of the participants had scores that were fifteen or more points different than that of the professionals in the field.

Table 4.06

Individual Differences: Lesson One

Range of Difference	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
0.00-4.99	n=3	n=5	n=3	n=5

5.00-9.99	n=3	n=3	n=2	n=2
10.00-14.99	n=1	n=1	n=1	n=2
15.00-19.99	n=3	n=1	n=3	n=1
20.00-24.99	n=0	n=0	n=1	n=0

(n)=number of students in that range

Teaching Lesson Two

Average Differences

The average self-assessment scores for teaching lesson two were less similar than those of teaching lesson one (Table 4.07). The lowest average was with Experimental Group III ($\bar{x}$ =83.73) while the highest average was Experimental Group I ($\bar{x}$ =91.456). The difference between the highest and lowest average was almost 8 points (*D*=7.726).

The average professional markings for teaching lesson two were closer together than the participants' self-assessments. The lowest average was in the Control Group ($\bar{x}$ =82.543) while the highest average was Experimental Group III ($\bar{x}$ =87.276). The difference between the highest and lowest score was *D*=4.733.

Differences between the self-assessments and the professional evaluations ranged from 0.362-5.547. Experimental Group I had the largest difference in scores (*d*=5.547), while Experimental Group II had the smallest difference (*d*=0.362).

Table 4.07

Lesson Two Average Evaluations for Each Group

Teaching Lesson	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
Two -Participant	85.092	91.456	85.093	84.73

-Professional	82.543	85.999	85.455	87.276
-Difference (*D*)	2.549	5.547	0.362	2.546

Individual Differences

Individual differences for lesson two and the number of participants who fell under each point range can be found in Table 4.08. Fourteen of the total participants scored within 5 points of the markings of the professionals in the field. Five of these fourteen participants had the exact same markings as that of the professionals. Three of the participants had scores that were fifteen or more points different from that of the professionals in the field.

Table 4.08

Individual Differences: Lesson Two

Range of Difference	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
0.00-4.99	n=4	n=2	n=5	n=3
5.00-9.99	n=5	n=5	n=1	n=3
10.00-14.99	n=0	n=3	n=3	n=3
15.00-19.99	n=0	n=0	n=1	n=1
20.00-24.99	n=0	n=0	n=0	n=0
25.00-29.99	n=1	n=0	n=0	n=0

(n)=number of students in that range

Teaching Lesson Three

Average Differences

The average self-assessment scores for teaching lesson three were similar. The lowest average was the Control Group ($\bar{x}$ =59.103), while the highest average was Experimental Group III ($\bar{x}$ =63.653). The difference between the highest and lowest average was 4.5 points (*D*=4.55).

The average professional markings for teaching lesson three were similar to one another (Table 4.09). The lowest average was in Experimental Group II ($\bar{x}$ =49.871), while the highest average was Experimental Group III ($\bar{x}$ =51.539). The difference between the highest and lowest score was *D*=1.668.

Differences between the self-assessments and the professional evaluations ranged from 8.396-12.31. Experimental Group I had the largest difference in scores (*d*=12.31) while the Control Group had the smallest difference (*d*=8.396).

Table 4.09

Lesson Three Average Evaluations for Each Group

Teaching Lesson	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
Three				
-Participant	59.103	62.436	61.218	63.653
-Professional	50.705	50.126	49.871	51.539
-Difference (*D*)	8.398	12.31	11.347	12.114

Individual Differences

Individual differences for lesson three and how many participants fell under each point range can be found in Table 4.10. Seven of the total participants scored within five points of the markings of the professionals in the field. One of these seven participants had the exact same markings as that of the professionals. Eleven of the participants had scores that were fifteen or more points different from that of the professionals in the field.

Table 4.10

Individual Differences: Lesson Three

Range of Difference	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
0.00-4.99	n=2	n=2	n=1	n=2
5.00-9.99	n=5	n=1	n=2	n=3
10.00-14.99	n=2	n=3	n=5	n=1
15.00-19.99	n=1	n=4	n=0	n=3
20.00-24.99	n=0	n=0	n=2	n=1

(n)=number of students in that range

Teaching Lesson Four

Average Differences

The average self-assessment scores for teaching lesson four ranged from ($\bar{x}$ =59.999) to ($\bar{x}$ =64.104). See Table 4.11. The lowest average was in Experimental Group I, while the highest average was Experimental Group III. The difference between the highest and lowest average was 4 points (*D*=4.105).

The average professional markings for teaching lesson four were similar. The lowest average was Experimental Group II ($\bar{x}$ =49.166), while the highest average was the Control Group ($\bar{x}$ =51.025). The difference between the highest and lowest score was *D*=1.859.

Differences between the self-assessments and the professional evaluations ranged from 10.513-13.334. Experimental Group III had the largest difference in scores (*d*=13.334), while the Experimental Group I had the smallest difference (*d*=10.513).

Table 4.11

Lesson Four Average Evaluations for Each Group

Teaching Lesson	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
Four				
-Participant	61.998	59.999	60.705	64.104
-Professional	51.025	49.486	49.166	50.77
-Difference (*D*)	10.973	10.513	11.539	13.334

Individual Differences

Individual differences for lesson four and how many participants fell under each point range can be found in Table 4.12. Five of the total participants scored within 5 points of the markings of the professionals in the field. One of these five participants had the exact same markings as that of the professionals. Twelve of the participants had scores that were fifteen or more points different from the markings of the professionals in the field.

Table 4.12

Individual Differences: Lesson Four

Range of Difference	Control Group	Experimental Group I	Experimental Group II	Experimental Group III
0.00-4.99	n=1	n=2	n=2	n=0
5.00-9.99	n=2	n=3	n=2	n=3
10.00-14.99	n=5	n=1	n=3	n=4
15.00-19.99	n=2	n=4	n=1	n=2
20.00-24.99	n=0	n=0	n=2	n=1

(n)=number of students in that range

Research Question Four

The analysis in this section provides information to answer the following research question: Do identifiable personality traits affect accuracy in self-assessment: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience? Two methods were utilized to analyze the data. The first was to check for homogeneity between the four groups and the second was to determine if any of the personality traits influenced participants' self-assessments.

The *Personality Profile* measured five categories: agreeableness, openness, extraversion, emotional stability, and conscientiousness. Average, high, and low scores for all the participants (N=40) can be found in Appendix R. Average, high, and low scores for the Control Group can be found in Appendix S, for Experimental Group I in Appendix T, for Experimental Group II in Appendix U, and for Experimental Group III in Appendix V.

To determine if the four groups were homogenous, data were analyzed using five, one-way analysis of variance (ANOVA) to compare the mean scores of each of the three experimental and one Control Group on the five personality traits. The agreeableness, extraversion, emotional stability, and conscientiousness personality traits were homogeneous amongst the four groups. Results for each group can be found in Table 4.13. Full data can be examined in Appendices W-Z.

Table 4.13

Personality Trait Data

Agreeableness	F (3, 36) = 2.06, p = 0.1230
Extraversion	F (3, 36) = 1.88, p = 0.1503
Emotional Stability	F (3, 36) = 0.33, p = 0.8007
Conscientiousness	F (3, 36) = 0.84, p = 0.4789

Significance, $p < .05$

Openness was the one personality trait for which a statistically significant difference was found, F (3, 36) = 3.11, p = 0.0384. The Control Group, Experimental Group II, and Experimental Group III were homogeneous, while Experimental Group I, Experimental Group II, and Experimental Group II were homogeneous. The Control Group and Experimental Group I were not homogeneous in the openness trait of the personality test. See Appendix AA for full data.

Once openness was found to be statistically different among the four groups, it was necessary to test for any influence on the self-assessment scores. Analysis of covariance (ANCOVA) procedures were utilized to test the data. The results of the test revealed that openness did not have an effect on the groups self-assessment scores, F (4,

35) = 0.57, p = 0.6835. Results are found in Table 4.14. Full data can be found in Appendix BB.

Table 4.14

ANCOVA Measuring Personality Traits to Group Scores

Source	Sum of squares	*df*	Mean sum of squares	F	Pr>F
Model	207.796	4	51.949	0.57	0.683
Error	3168.9226	35	90.5406		
Corrected Total	3376.7195	39			

Summary of Chapter

Participants (N=40) in this study were enrolled in one of two classes, either one for music education majors or one for education majors in other fields. All students were undergraduates at a large public university in the southeastern United States. Participants taught four lessons with a self-assessment completed immediately after each teaching episode. The Experimental Groups either watched a videotape of their teaching, examined peer critiques, or a combination of both videotape analysis and peer critique while completing their self-assessments. Both quantitative and qualitative measures were used to categorize and analyze the data.

Quantitative measures were utilized to test if videotape, peer critique or a combination of both videotape and peer critique influenced an individual's ability to self-assess. There was no statistically significant difference in self-assessment scores among the four groups.

Participant responses to posed questions revealed that the use of videotape did allow the individual to find teaching weaknesses at a quicker rate. Responses to peer critique were mixed: some preferred to use it while others stated that it was often unreliable and required too much time to examine.

In order to see if individual self-assessments matched those of the professionals, simple quantitative measures were used. Eighteen of the forty participants were within 5 points of the professional markings for lesson one. Fourteen participants were within 5 points of the professional markings for lesson two, only seven for lesson three, and just five for lesson four. For the majority of the participants, scores did not match those of the professionals in the field.

Multiple quantitative measures were utilized to test the effects of individual personality traits on participant self-assessments. Agreeableness, extraversion, emotional stability, and conscientiousness were homogeneous among all four groups. Openness was significantly different between the Control Group and Experimental Group I. Openness was then compared to participants' self-assessment scores. It was determined that openness did not affect the self-assessment markings.

CHAPTER 5

DISCUSSION OF RESULTS AND STUDY IMPLICATIONS

The motivation for this study was a direct result of observing practicum students in the field struggle with their self-assessment skills. Since evaluation is a continual part of the practicing teacher's routine, it seemed appropriate to examine techniques that could possibly lead to better self-assessments.

In order to produce quality teachers, higher education institutes should provide preservice teachers with the skills needed to connect what they learn (theory) with actual practice. Part of making this connection is teaching them how to properly self-assess. Self-assessment "converts action that is merely appetitive, blind, and impulsive into intelligent action" (Dewey, 1933, p. 17). This study attempted to address these issues by examining how participants self-assess and if videotape, peer critique, or a combination of both videotape and peer critique help the individual to more accurately self-assess. Four specific questions were addressed.

Research Question One

Does the use of videotape analysis and a peer critique process result in more accurate self-assessment?

Analyzed data revealed that videotape analysis and peer critique did not seem to enable the preservice teacher to more accurately assess their teaching. The participants' self-assessment markings were compared to that of the professionals in the field for each of the four teaching lessons. For all four groups, control and three experimental, average

scores grew further away from the markings of the professionals. These results match the results of Kpanja (2001) and Perlberg (1983), who noted little change when videotape was used, and Topping (1998), who described the inaccuracy of peer critiques. Several factors could have contributed to these results: different self-assessment forms, study fatigue, and outside factors.

Different Forms

Each of the four teaching lessons utilized a different evaluation form (Appendices D-G). All of the forms were assigned a Likert-type scale for scoring: form one had a total of eighty-eight points, form two a total of fifty-five points, form three a total of 156 points, and form four a total of 156 points. Since forms three and four contained almost twice the number of possible points as form one and three times the points of form two, it is likely that this resulted in the higher averages following lessons three and four. With more choices for the participant, the likelihood of greater variance increases.

Participant Fatigue

During the debriefing, a few of the participants mentioned that as the study progressed they spent less time analyzing their teaching. This included less time reflecting on their teaching, less time viewing their videotape, and less attention given to the markings on the peer critiques. When asked about the cause, several participants responded. All of the participants in Experimental Groups I and III agreed that the viewing of their teaching on videotape was helpful. The problem surfaced with the viewing time. Although they liked the option to rewind and continually review their lessons, they had a problem with the amount of time that it took to view their videotapes.

A few of the participants admitted spending less time viewing their videotapes as the study progressed.

Experimental Group II and Experimental Group III examined evaluations from their peers while self-assessing. Several of the participants in these two groups mentioned that some of the forms seemed arbitrarily marked. Others discredited the forms, stating that one peer evaluator would mark that the participant completed a task while another peer evaluator would mark that the participant did not complete the task. These inaccuracies are in line with the findings of Chaves, Baker, and Chaves (2006), who noted that peer critiques were often less reliable than the self-assessment.

Outside Factors

Factors outside the control of the researcher could have influenced the results of the self-assessment forms. Participants commented very often on the stresses of the semester. Other classes, job obligations, and other personal circumstances could have taken away some of the focus on the study. When individuals become overwhelmed they often prioritize their schedules; unfortunately, this research study may not have been the most important.

Research Question Two

Does the practice of continual video analysis and peer critique result in the pre-service teacher being able to identify teaching weaknesses at a quicker rate?

Qualitative procedures were used to answer this question. Results indicated that videotape analysis did enable participants to identify teaching weaknesses at a quicker rate than those participants in the Control Group. The participants in the videotape analysis groups, Experimental Groups I and III, stated that the videotape allowed them to

easily see discrepancies in their teaching. By watching the videotape of their lessons multiple times, they were able to focus on different teaching traits each time they viewed the video. Several of the participants stated that it was easy to identify weaknesses when watching the videotape. These statements confirm the findings of Hourigan (2006) that videotape observation and reflection help individuals evaluate themselves.

Teachers, somewhat like ensemble conductors, are required to consistently multi-task. Self-assessment as another layer to this list can be extremely difficult. The use of videotape recording relieves the individual from having to focus on the evaluation while they are teaching.

Results for individual participants in the peer critique Experimental Groups yielded mixed results. Nine of the thirteen participants utilizing peer critiques remarked that the evaluations did help them to more quickly realize teaching weaknesses. This agrees with the findings of Chaves, Baker, and Chaves (2006) and Topping (1998). The other four participants that felt that the peer critiques were not effective in helping them identify teaching discrepancies at a quicker rate all stated that the markings could not be trusted. They each felt that the peer marks presented differing views and that the peers could not accurately critique the participants teaching. Other comments indicated that it was too time consuming to sort through a large stack of peer critiques. This agrees with Falchikov and Goldfinch (2000), who recommend that a smaller number of peers be placed in each group rather than a larger number. One participant recommended verbal rather than written feedback. The potential problem with verbal feedback is maintaining objectivity since the teacher can easily identify who is making the comment.

Although there was some negative feedback for the videotape analysis and peer critique process, it did enable participants to identify teaching weaknesses more easily and quickly than those participants who did not use these techniques.

Research Question Three

To what extent do the self-assessments agree with those of experts?

Quantitative results comparing the self-assessments to those of the professionals reveal that most participants' self-assessment scores did not match those of the professionals in the field. Of the forty participants, eighteen scored within 5 points of the professionals on their self-assessments following teaching lesson one, only fourteen for lesson two, seven for lesson three, and five for lesson four. Participant markings matched exactly those of the professionals only eleven times during the course of the project: 4 times during lesson one, 5 times during lesson two, and 1 time during lessons three and four.

There are several reasons that the participants and professional evaluations did not match. These include: a lack of training in completing the self-assessments, not understanding how the professionals in the field are evaluating, having no baseline data, varying total points between the evaluation forms, and experience.

Lack of Training

When the participants began the study there was no training in accurately completing the self-assessment forms. There was no baseline data or standards to compare the forms to. This could have easily caused participants to mark either higher or lower than they should have. The result would have been scores that did not match those of the professionals in the field.

Professionals Evaluation

Another factor that could have resulted in scores that did not match those of the professionals in the field was a lack of knowledge of how the professionals in the field were evaluating the teaching episodes. Although the forms contained specific questions, how those questions were interpreted could possibly mean one thing to the professionals and another to the participants. Having the professionals explain to the participants how they were evaluating and clearing up any confusion with the four forms questions might have eliminated this factor. Although as Kottkam, Provenzo, and Conn (1986) found, the individual may dismiss the supervisor's feedback as not being useful.

Baseline Data

Any time individuals are asked to judge or rank order something, many differing results will usually emerge. One possible way to control this is to provide baseline data to the participants before beginning the research study. The participants could either teach a lesson and then talk through the self-assessment with the professionals in the field or they could watch a video or another participant teaching a lesson and then talk through the evaluation process. Once baseline data or standards were established and understood by all of the participants, the results of each self-assessment might better resemble those of the professionals.

Varying Points Between Forms

A fourth factor that could contribute to the variance between participant and professional markings is the changing number of points between forms. When examining the data, participants' scores were closer to the professionals after teaching lessons one

and two. Nine of the eleven scores that matched exactly those of the professionals in the field came from evaluations following lessons one and two.

Lessons one and two contained fewer than 100 possible points, while lessons three and four contained more than 150 points. It would seem that with more options, scores would be further apart. Even though the points were converted into percentages to make the comparisons from lesson to lesson, it would seem appropriate to have forms that contained the same number of questions and possible markings. Had the forms followed this method, the participants and professional scores may have improved with each passing lesson.

Experience

The final factor that could have created the disparity between participants and professional markings is the level of experience. All of the participants had just begun their teacher training. The members of the professional panel had been teaching for years and had experience across all levels of schooling. Comparing experienced professionals to beginning preservice teachers could easily cause great variance in the evaluation marks. Although an easy identifiable factor, it would be difficult to account for.

Even though the participants' scores were further from the markings of the professionals in the field, it is interesting to note that the standard deviation was smaller with each passing lesson. This indicates that each participant's score became more consistent within their respective research group. The only explanation for this is that as the two classes progressed, the participants learned what to focus on.

Research Question Four

Do identifiable personality traits affect accuracy in self-assessment: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience?

The personality inventory tested five traits: agreeableness, openness, extraversion, emotional stability, and contentiousness. With the exception of openness, all traits were homogeneous among the four groups. Openness was significantly different between the Control Group and Experimental Group I. When compared to the self-assessment scores it was found that openness did not affect them.

Although there was no affect found in this study, it is very probable that personality could influence an individual's self-assessment markings. Individuals who are more introverted than extroverted may be more critical of themselves. Bergee and Cecconi-Roberts (2002), Cheney (2005), Fuller and Manning (1973), Lethco (1999), and Perlberg (1983) all note that individuals seeking self-fulfillment are usually eager to identify any behaviors that are weak and need modification.

Very open individuals may complete their self-assessments with greater accuracy and truth than someone who is private and less open. An individual who is emotionally stable may have self-assessments that are more consistent than someone whose emotions are in constant fluctuation. As Fuller and Manning (1973) state, individuals who have high self-esteem and a good image of themselves are most likely to benefit from self-assessment.

Conclusions

1. Although the inclusion of a videotape analysis and peer critique process has proved successful in allowing preservice teachers to more accurately assess their teaching, it was not statistically significant in this study.

2. The use of both videotape analysis and a peer critique process was found to enable participants to identify teaching weaknesses at a quicker rate than those who just self-assessed. Use of the videotape was found to be the more useful and accurate than the peer critiques.

3. Only eleven times during the course of the study did the participants evaluations exactly match those of the professionals in the field. Participants' self-assessments were within five points of the professionals a total of forty-four times during the course of the study. Overall, the participants' self-assessments did not match those of the professionals in the field.

4. With the exception of openness, all of the personality traits were homogeneous among the four groups and thus did not affect the self-assessment scores of the groups. When tested, the one significantly different trait, openness, was found to have no impact on participants' self-assessments.

Implications for Education

Evaluation and assessment is continually emphasized at all levels of education. Students, teachers, and administrators should be taught what to expect on an evaluation

and how to properly self-assess. Although a challenging task, higher education institutions that train teachers should develop curriculum that teaches preservice teachers how to properly self-assess. Curriculum should focus on teaching techniques, methods, verbal behaviors, and non-verbal behaviors. Preservice teachers should have the opportunity to watch video of both good and bad teaching episodes and to develop their evaluation skills. They should be given multiple opportunities to teach with feedback immediately following. They should have the opportunity to talk one-on-one with professionals in their fields. Ideas and ways to correct discrepancies in their teaching should be continually presented.

This study was intended to determine the benefits of using videotape analysis and peer critiques to improve the accuracy of individual self-assessments. Although there were no significant differences between groups, videotape analysis and peer critique should be used as supplemental devices in the training of preservice teachers.

If all teacher training institutions would adapt a curriculum that emphasizes self-assessment and self-assessment techniques, it would be possible to produce a new group of practicing teachers who constantly monitor their own teaching. If they monitor their own teaching, they are more likely to have successful students, up-to-date teaching practices, and higher evaluation marks.

Implications for Future Research

Future research should continue to investigate self-assessment and the preservice teacher. Although there is a plethora of evaluation research, a meta-analysis yields many differing results. Expanded research based on this study should consider eliminating the limitations found herein. The first consideration is the sample size. This study contained

forty participants. Although it included almost all of the students in the two classes, it was a relatively small sample size for general inferences to be made. A sample size of 100 or greater might reveal data that is more consistent among the Control and three Experimental Groups. A larger sample size may be more representative of the overall population and thus result in a greater acceptance of any significant results.

A second consideration is the time frame. The experimental portion of this project lasted fifteen weeks or one semester. An on-going study that last several years might reveal data with greater significance. This could be accomplished by following a group of preservice participants through their entire teacher education curriculum.

A third consideration is the geographic location. This study was completed at a large public university in the southeastern United States. A study with participants from a variety of geographic regions, public and private institutions, and large and small populations would seem appropriate.

The two classes consisted of an undergraduate music education majors class and an undergraduate music class for education majors of other fields. Study of preservice teachers in classes outside of music education might reveal differing results. Other areas could include preservice teachers majoring in: special education, foreign language, elementary education, secondary education, or the other fine arts. Other research could include participants representing multiple education fields in one study.

Future research on self-assessment should explore changes to the higher education curriculum. If training preservice teachers to properly self-assess is of high priority, then it would seem necessary to include it in the curriculum. Questions to consider are: 1) Should self-assessment be a small portion of certain classes? 2) Should it be a stand-alone

class? 3) How many semesters or years of self-assessment training are necessary to thoroughly teach students to properly self-evaluate?

Another limitation of this study was the variables. The two areas influencing self-assessments that were examined were videotape analysis and peer critiques. Future research could isolate these areas, studying the effects of videotape analysis on self-assessment solely or only peer critiques effects on self-assessments.

Other factors influencing self-assessment, such as portfolios, supervisor evaluations, and journaling, should be examined. These techniques could be isolated or combined as various independent variables.

Although personality did not seem to influence participants' self-assessments, it might be important to include a personality test in future research. Each category of the *Internet Personality Profile Survey*—agreeableness, openness, extraversion, emotional stability, and conscientiousness—could be isolated or tested as groups against an individual's self-assessment.

REFERENCES

Abbott, W. P. (1965). *An experimental study of self-confrontation in counseling*. Dissertations Abstracts International. University of Southern California, Los Angeles.

Airasian, P. W. (1995). *Teacher self-evaluation: The literature in perspective*. Kalamazoo, MI: Center for Research on Educational Accountability and Teacher Evaluation at Western Michigan University.

Albert, C. R., & Hipp, R. (1976). Videotape recording in the preparation of college sociology teachers. *Teaching Sociology, 3*(3), 327-338.

Anderson, J. B., & Freiberg, H. J. (1995). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly* (Winter), 77-91.

Astin, A. W. (1975). *Preventing students form dropping out*. San Francisco: Jossey-Bass.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 25*(4), 297-308.

Attinello, J. R., & Lare, D. (2006). The value of teacher portfolios for evaluation and professional growth. *NASSP Bulletin, 90*(2), 132-152.

Bailey, G. D. (1981). *Teacher self-assessment: A means for improving classroom instruction*. Washington, D.C.: National Education Association.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, New York: Prentice-Hall, Inc.

Barber, L. W., & Lewis, S. (1986). *Improving your own instruction: Self-assessment and peer review*. Bloomington, Indiana: Phi Delta Kappa, Center on Evaluation, Development and Research.

Barnes, G. L. V. (1998). *A comparison of self-efficacy and teaching effectiveness in preservice string teachers.* Unpublished doctoral dissertation, The Ohio State University, Columbus.

Barrett, K. R. (1983). A hypothetical model of observing as a teaching skill. *Journal of Teaching in Physical Education, 3*(1), 22-31.

Beck, R. J., King, A., & Marshall, S. K. (2002). Effects of video case construction on preservice teachers' observations of teaching. *The Journal of Experimental Education, 70*(4), 345-361.

Bedics, R. A., & Webb, J. N. (1971). Measuring the self-evaluation of teaching behavior through the use of video tape. Florida. (ERIC Document Reproduction Service No. ED051079).

Benne, K. D., Bradford, L. P., & Lippitt, R. (1964). The laboratory method. In L. P. Bradford, J. R. Gibb & K. D. Benne (Eds.), *T-group theory and laboratory method*. New York: John Wiley.

Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education, 41*(1), 19-27.

Bergee, M. J., & Cecconi-Roberts, L. (2002). Effects of small-group peer interaction on self-evaluation of music performance. *Journal of Research in Music Education, 50*(3), 256-268.

Bernstein, D. J., Jonson, J., & Smith, K. (2000). An examination of the implementation of peer review of teaching. In K. E. Ryan (Ed.), *Evaluating teaching in higher education: a vision for the future*. San Francisco: Jossey-Bass.

Berube, B., & Dexter, R. (2006). Supervision, evaluation and NCLB: Maintaining a most highly qualified staff. *Catalyst for Change, 34*(2), 11-17.

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-147.

Blanton, L. P., Sindelar, P. T., & Correa, V. I. (2006). Models and measures of beginning teacher quality. *The Journal of Special Education, 40*(2), 115-127.

Bloom, J. (1969). Videotape and the vitalization of teaching. *Journal of Teacher Education, 20*(3), 311-315.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529-549.

Bullough, R. (1989). *First-year teacher: a case study*. New York: Teachers College Press.

Burrack, F. W. (2001). *Using reflection and video self-confrontation to uncover the instructional thought development of student teachers in instrumental music.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign.

Butler, A. (1999). *The relationship among preservice music education teachers' conceptions of teaching effectiveness, microteaching experiences, and teaching performance.* Unpublished doctoral dissertation., The University of Arizona, Tucson.

Carr, R. A. (1977). The effects of specific guidelines on the accuracy of student self-evaluation. *Canadian Journal of Education / Revue canadienne de l'Ãducation, 2*(4), 65-77.

Carter, K. S., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education, 39*(3), 25-31.

Cassidy, J. W. (1990). Effect of intensity training on preservice teachers' instruction accuracy and delivery effectiveness. *Journal of Research in Music Education, 38*(3), 164-174.

Cassidy, J. W. (2005). The effect of focus of attention and teaching experiences on perceptions of teaching effectiveness and student learning. *Journal of Research in Music Education, 53*(3), 222-233.

Chaves, J. F., Baker, C. M., & Chaves, J. A. (2006). Self, peer, and tutor assessments of MSN competencies using the PBL-evaluator. *Journal of Nursing Education, 45*(1), 25-31.

Chen, W., Burry-Stock, J. A., & Rovengno, I. (2000). Self-evaluation of expertise in teaching elementary physical education from constructivist perspectives. *Journal of Personnel Evaluation in Education, 14*(1), 25-45.

Cheney, G. (2005). A look at the MTNA assessment tools for the independent music teacher. *The American Music Teacher, 55*(1), 86.

Clark, C. M. (1988). Asking the right questions about teacher preparation: Contributions of research on teacher thinking. *Educational Researcher, 17*(2), 5-12.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Clarke, S. M. (1958). Opinions and practices of supervisors of student teachers in selected music schools. *Journal of Research in Music Education, 6*(1), 62-67.

Claudson, W. D. (1969). The philosophy of Julia E. Crane and the origin of music teacher training. *Journal of Research in Music Education, 17*(4), 399-404.

Colwell, C. M. (1993). *Effect of self-evaluation and teaching setting on teacher intensity behaviors among preservice elementary education majors enrolled in a music methods course.* Unpublished doctoral dissertation, Louisiana State University and Agricultural & Mechanical College, Baton Rouge.

Colwell, C. M. (1995). Effect of teaching setting and self-evaluation on teacher intensity behaviors. *Journal of Research in Music Education, 43*(1), 6-21.

Combs, A. W. (1964). The personal approach to good teaching. *Educational Leadership* (3), 369-399.

Darling-Hammond, L. (1988). Accountability and teacher professionalism. *American Educator, 12*(4), 8-13, 38-43.

DeBacy, D. L. (1969). *The effect of viewing videotapes of a selected sport skill performed by self and others on self-assessment.* Ohio State University, Ann Arbor, Michigan.

Dewey, J. (1933). *How we think, a restatement of the relation of reflective thinking to the educative process*. Boston, New York: D.C. Heath and company.

DiGangi, S. A., Maag, J. W., & Rutherford, R. B., Jr. (1991). Self-graphing of on-task behavior: Enhancing the reactive effects of self-monitoring on on-task behavior and academic performance. *Learning Disability Quarterly, 14*(3), 221-230.

Donnelly, A. M. (2005). Let me show you my portfolio! Demonstrating competence through peer interviews. *Action in Teacher Education, 27*(3), 55-63.

Duke, R. A., & Blackman, M. D. (1991). The relationship between observers' recorded teacher behavior and evaluation of music instruction. *Journal of Research in Music Education, 39*(4), 290-297.

Duke, R. A., & Prickett, C. A. (1987). The effect of differentially focused observation on evaluation of instruction. *Journal of Research in Music Education, 35*(1), 27-37.

Eckrich, J., & Widule, C. J. (1994). The effects of video observational training on video and live observation proficiency. *Journal of Teaching in Physical Education, 13*, 216-237.

Evaluating teaching. (1963). *College Composition and Communication, 14*(3), 191-192.

Falchikov, N. (1999). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*, 175-187.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322.

Fendler, L. (2003). Teacher reflection in a hall of mirrors: Historical influences and political reverberations *Educational Researcher, 32*(3), 16-25.

Fletcher, A. (2005). *Meaningful student involvement: guide to students as partners in school change*. Olympia, WA: Common Action.

Frager, A. M. (1985). Video technology and teacher training: A research perspective. *Educational Technology, 25*(7), 20-22.

Fuller, F. F., & Manning, B. A. (1973). Self-confrontation reviewed: A conceptualization for video playback in teacher education. *Review of Educational Research, 43*(4), 469-528.

Garmston, R. (1987). How administrators support peer coaching. *Educational Leadership, 44*(5), 18-26.

Geertsma, R. H., & Reivich, R. S. (1965). Repetitive self-observation by videotape playback. *Journal of Nervous and Mental Disease, 141*(1), 29-41.

Gillis, A. J. (2001). Journal writing in health education. *New Directions for Adult and Continuing Education, 90*, 49-58.

Goldstein, J., & Noguera, P. A. (2006). A Thoughtful approach to teacher evaluation. *Educational Leadership, 63*(6), 31-37.

Gonzo, C., & Forsythe, J. (1976). Developing and using videotapes to teach rehearsal techniques and principles. *Journal of Research in Music Education, 24*(1), 32-41.

Gordon, S. P. (2002). *Professional development for school improvement: Empowering learning communities*. Boston: Allyn & Bacon.

Gunter, P. L., & Reed, T. M. (1996). Self-evaluation of instruction: a protocol for functional assessment of teaching behavior. *Intervention in School and Clinic, 31*, 225-230.

Guthrie, J. W. (2005). An "Education professions performance development act": A prospectus for providing "highly qualified" and more motivated teachers and leaders for America's schools. *Peabody Journal of Education, 80*(3), 6-14.

Hamann, D. L., Baker, D. S., McAllister, P. A., & Bauer, W. I. (2000). Factors affecting university music students' perceptions of lesson quality and teaching effectiveness. *Journal of Research in Music Education, 48*(2), 102-113.

Hamann, D. L., Lineburgh, N., & Paul, S. (1998). Teaching effectiveness and social skill development. *Journal of Research in Music Education, 46*(1), 87-101.

Hargreaves, A. (1995). Renewal in the age of paradox. *Educational Leadership, 52*(7), 14-19.

Hirschfeld, A. G. (1968). Videotape recordings for self-analysis in the speech classroom. *Speech Teacher, 17*, 116-118.

Hollingsworth, S. (1989). Prior beliefs and cognitive change in learning to teach. *American Educational Research Journal, 26*(2), 160-189.

Holzman, P. S. (1969). On hearing and seeing oneself. *Journal of Nervous and Mental disease, 148*(3), 198-209.

Hourigan, R. (2006). The use of the case method to promote reflective thinking in Music teacher education. *Update - Applications of Research in Music Education, 24*(2), 33-44.

Irvine, J. (1983). The accuracy of pre-service teachers' assessment of their classroom behaviors. *Journal of Research and Development in Education, 17*, 25-31.

Irwin, D. M., & Bushnell, M. M. (1980). *Observation strategies for child study*. New York: Holt, Rinehart & Winston.

Karasar, N. (1970). *Impact of video feedback on teachers' eye-contact mannerisms in microteaching.* Unpublished doctoral dissertation, Ohio State University, Ann Arbor.

Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague assessment. *Journal of Personnel Evaluation in Education, 14*(1), 67-87.

Keller, C. L., Brady, M. P., & Taylor, R. L. (2005). Using self evaluation to improve student teacher interns' use of specific praise. *Education and Training in Developmental Disabilities, 40*(4), 368-376.

Kerrins, J. A., & Cushing, K. S. (2000). Taking a second look: Expert and novice differences when observing the same classroom teaching segment a second time. *Journal of Personnel Evaluation in Education, 14*(1), 5-24.

Kostka, M. J. (1997). Effects of self-assessment and successive approximations on "knowing" and "valuing" selected keyboard skills. *Journal of Research in Music Education, 45*(2), 273-281.

Kottkamp, R. B., Provenzo, E. F., & Conn, M. M. (1986). Stability and change in a profession: two decades of teacher attitudes, 1964-1984. *Phi Delta Kappan, 47*(6), 559-567.

Kpanja, E. (2001). A study of the effects of video tape recording in microteaching training. *British Journal of Educational Technology, 32*(4), 483-486.

Kremer-Hayon, L. (1993). *Teacher self-evaluation: teachers in their own mirror*. Boston: Kluwer Academic.

Kretchmer, D. L. (2002). *Developing pre-service teacher self-efficacy to integrate music in elementary classrooms: An investigation in growth through participation, observation and reflection.* Unpublished doctoral dissertation, University of Colorado, Boulder.

Kuhn, D. (1991). *The skill of argument*. New York: Cambridge University Press.

Lepard, D. H. (2002). Using peers and technology to strengthen leadership. *Journal of Personnel Evaluation in Education, 16*(1), 11-28.

Lethco, L.-A. M. (1999). *Preparing undergraduate music majors to teach beginning instrumentalists: The effects of self-evaluation, teacher observation, and performance-oriented instructional approaches on teacher behaviors and pupil responses.* Unpublished doctoral dissertation, Louisiana State University and Agricultural & Mechanical College, Baton Rouge.

Lieberman, A. (1988). Teachers and principals: turf, tension, and new tasks. *Phi Delta Kappan, 69*(9), 648-653.

Liu, E. Z.-f., Lin, S. S. J., & Yuan, S. M. (2002). Alternatives to instructor assessment: A case study of comparing self and peer assessment with instructor assessment under a networked innovative assessment procedures. *International Journal of Instructional Media, 29*(4), 395-404.

Logue, P. E., Zenner, M., & Gohman, G. (1968). Video-tape role playing in the job interview. *Journal of Counseling Psychology, 15*(5), 436-438.

Lunenberg, M., & Willemse, M. (2006). Research and professional development of teacher educators. *European Journal of Teacher Education, 29*(1), 81-98.

Madsen, K. (2003). The effect of accuracy of instruction, teacher delivery, and student attentiveness on musicians' evaluation of teacher effectiveness. *Journal of Research in Music Education, 51*(1), 38-50.

Manfredo, J. (2006). Effective time management in ensemble rehearsals. *Music Educators Journal, 93*(2), 42-46.

Maranzano, C. (2000). Music teacher performance evaluation: A call for more inclusive models. *Journal of Personnel Evaluation in Education, 14*(3), 267-274.

Marion, R. F., & Samaniego, F. A. (1981). The supervision of teaching assistants: A new use of videotape. *The Modern Language Journal, 65*(3), 273-280.

Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86*(10), 727-735.

Marshall, K. (2008). Is supervising the heck out of teachers the answer? *Education Week, 27*(36), 23, 25.

Martin, K. M. (2004). Assuring novice teacher competency in Ohio and Japan. *Action in Teacher Education, 26*(3), 52-58.

Merkley, D., & Hoy, M. P. (1985). Teacher-on-television: A new mode of preservice classroom observation. *Phi Delta Kappan, 66*(5), 373-374.

Morse, K. R., Kysilka, M., & Davis, O. L. (1970). *Effects of different types of supervisory feedback on teacher candidates' development of refocusing behaviors*. Austin, TX: The University of Texas at Austin.

Natriello, G. (1984). Problems in the evaluation of students and student disengagement from secondary schools. *Journal of Research and Development in Education, 17*, 14.

Nitz, D. A. (1982). "Student, know thyself" through values clarification. *Music Educators Journal, 69*(4), 44-47.

Olivero, J. L. (1966). *The use of video recordings in teacher education*. Stanford: College of Education, Stanford University Press.

Ory, J. C. (2000). Teaching evaluation: past, present, future. In K. E. Ryan (Ed.), *Evaluating teaching in higher education: a vision for the future*. San Francisco: Jossey-Bass.

Osterman, K. F., & Kottkamp, R. B. (1993). *Reflective practice for educators: improving school through professional development*. Newbury Park, CA: Corwin Press, Inc.

Ovando, M. (2001). Teachers' perceptions of a learner-centered teacher evaluation system. *Journal of Personnel Evaluation in Education, 15*(3), 213-231.

Pailliotet, A. W. (1995). I never saw that before: A deeper view of video analysis in teacher education. *Teacher Educator, 31*(2), 138-156.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.

Panhorst, D. L. (1971). Current practices in the evaluation of student teachers in music. *Journal of Research in Music Education, 19*(2), 204-208.

Pease, J. J. (1972). *The effectiveness of video tape feedback on driving performance and self-evaluation.* Unpublished doctoral dissertation, University of Wisconsin, Madison.

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*(2), 215-253.

Perlberg, A. (1983). When professors confront themselves: Towards a theoretical conceptualization of video self-confrontation in higher education. *Higher Education, 12*(6), 633-663.

Perlberg, A., Peri, J. N., Weinreb, M., Nitzan, E., Shimron, J., & O'Bryant, D. (1971). *Studies on the use of video-tape recordings and microteaching techniques to improve university teaching*. Paper presented at the American Educational Research Association.

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79.

Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices.* Thousand Oaks, CA: Corwin Press.

Peterson, K. D., Kelly, P., & Caskey, M. (2002). Ethical considerations for teachers in the evaluation of other teachers. *Journal of Personal Evaluation in Education. 16, 4*(317-324).

Peterson, K. D., Stevens, D., & Mack, C. (2001). Presenting complex teacher evaluation data: Advantages of dossier organization techniques over portfolios. *Journal of Personnel Evaluation in Education, 15*(2), 121-133.

Phillips, J. M. (2005). Strategies for active learning in online continuing education. *The Journal of Continuing Education in Nursing, 36*(2), 77-83.

Price, H. E. (1992). Sequential patterns of music instruction and learning to use them. *Journal of Research in Music Education, 40*(1), 14-29.

Roth, W. M., & Tobin, K. (2000). The implications of coteaching/cogenerative dialogue for teacher evaluation: Learning from multiple perspectives of everyday practice. *Journal of Personnel Evaluation in Education, 15*(1), 7-29.

Sawyer, L. (2001). Revamping a teacher evaluation system. *Educational Leadership, 58*(5), 44-47.

Schmidt, M. E. (1994). *Learning from experience: Influences on music student teachers' perceptions and practices.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Schroeder, M. L. (1996). Lesson design and reflection. *Mathematics Teaching in the Middle School, 1*(18), 648-652.

Schueler, H., & Gold, M. J. (1964). Video recordings of student teacher: A report of the hunter college research project evaluating the use of kinescopes in preparing student teachers. *Journal of Teacher Education, 15*, 358-364.

Schulz, R. (2005). The practicum: More than practice. *Canadian Journal of Education / Revue canadienne de l'Ãducation, 28*(1/2), 147-167.

Scriven, M., Stodolsky., Good., Mulryan., Glass., Bird., et al. (1990). Teacher selection. In J. Millman & J. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: assessing elementary and secondary school teachers*. Newury Park, CA: Corwin Press, Inc.

Shedd, J. B. (1985). *From the front of the classroom: a study of the work of teachers*. Ithaca, NY: Organizational Analysis and Practice, Inc.

Shepard, L. A. (1979). Self-acceptance: The evaluative component of the self-concept construct. *American Educational Research Journal, 16*(2), 139-160.

Sherin, M. G. (2000). Viewing teaching on videotape. *Educational Leadership, 57*(8), 36-38.

Simpson, R. H. (1964). Music instructors' use of self-evaluation tools. *Music Educators Journal, 50*(5), 82-88.

Simpson, R. H. (1966). *Teacher self-evaluation*. New York: Macmillan.

Smith, E. B., & Kight, S. S. (1959). Effects of feedback on insight and problem solving efficiency in training groups. *Journal of Applied Psychology, 43*(3), 209-211.

Stegman, S. F. (2007). An exploration of reflective dialogue between student teachers in music and their cooperating teachers. *Journal of Research in Music Education, 55*(1), 65-82.

Sterling, D. R. (2008). Assessing student presentations from 3 perspectives. *Science Scope, 31*(5), 34-37.

Steward, M. S., & Steward, D. S. (1970). Teacher, teach yourself. *Audiovisual Instruction, 15*, 26-27.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher, 13*(9), 11-19.

Stoll, C. S. (1973). Videotape as an aid in sociology instruction. *Teaching Sociology, 1*(1), 38-47.

Storeygard, J. F. (1995). Reflection on video: One teacher's story. *Journal of Staff Development, 16*(3), 25-29.

Stout, C. J. (1989). Teachers' views of the emphasis on reflective teaching skills during their student teaching. *The Elementary School Journal, 89*(4), 511-527.

Stroh, T. F. (1968). *Videotape feedback in the development of listening skills by industrial salesmen.* Unpublished doctoral dissertation, Columbia University, New York.

Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. New York: Eye on Education.

Stuart, M. (1979). The use of videotape recordings to increase teacher trainees' error detection skills. *Journal of Research in Music Education, 27*(1), 14-19.

Sutherland, K. S., & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms for students with emotional and behavioral disorders. *The Journal of Special Education, 35*(3), 161-171.

Taebel, D. K. (1992). The evaluation of music teachers and teaching. In R. Colwell (Ed.), *Handbook of research on music teaching and learning* (pp. 310-329). New York, NY: G. Schirmer Inc.

Taebel, D. K. (1980). Public school music teachers' perceptions of the effect of certain competencies on pupil learning. *Journal of Research in Music Education, 28*(3), 158-197.

Thompson, L. K. (2007). Considering beliefs in learning to teach music. *Music Educators Journal, 93*(3), 30-35.

Tjornehoj, K. A. S. (2001). *An investigation into the use and effectiveness of video modeling of conducting for pre-service music educators.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis - Saint Paul.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

Turrentine, C. G. (2001). A comparison of self-assessment and peer assessment of leadership skills. *NASPA Journal, 38*(3), 361-371.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

Verrastro, R. E. (1975). Verbal behavior analysis as a supervisory technique with student teachers of music. *Journal of Research in Music Education, 23*(3), 171-185.

Wagner, K. (2006). Benefits of reflective practice. *Leadership, 36*(2), 30-32.

Walls, R. T., Nardi, A. H., & Von Minden, A. M. (2002). The characteristics of effective and ineffective teachers. *Teacher Education Quarterly, 29*(1), 39-48.

Ward, M. A. (2005). An empirical assessment of peer review reliability: The positive instrument paradox. *Journal of Education for Business, 81*(1), 35-40.

Webster, N. (1953). *New collegiate dictionary. A Merriam-Webster*. Springfield, MA: G. & C. Merriam Co.

Wilkerson, D. J., Manatt, R. P., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 feedback for teacher evaluation. *Journal of Personnel Evaluation in Education, 14*(2), 179-192.

Winn, W. (1974). Videotaping teaching practice: Strengths and weaknesses. *Audiovisual Instruction, 19*(1), 18, 20.

Worthy, M. D. (2005). The effects of self-evaluation on the timing of teacher and student behaviors in lab rehearsals. *Journal of Music Teacher Education, 15*(1), 8-14.

Yarbrough, C. (1987). The relationship of behavioral self-assessment to the achievement of basic conducting skills. *Journal of Research in Music Education, 35*(3), 183-189.

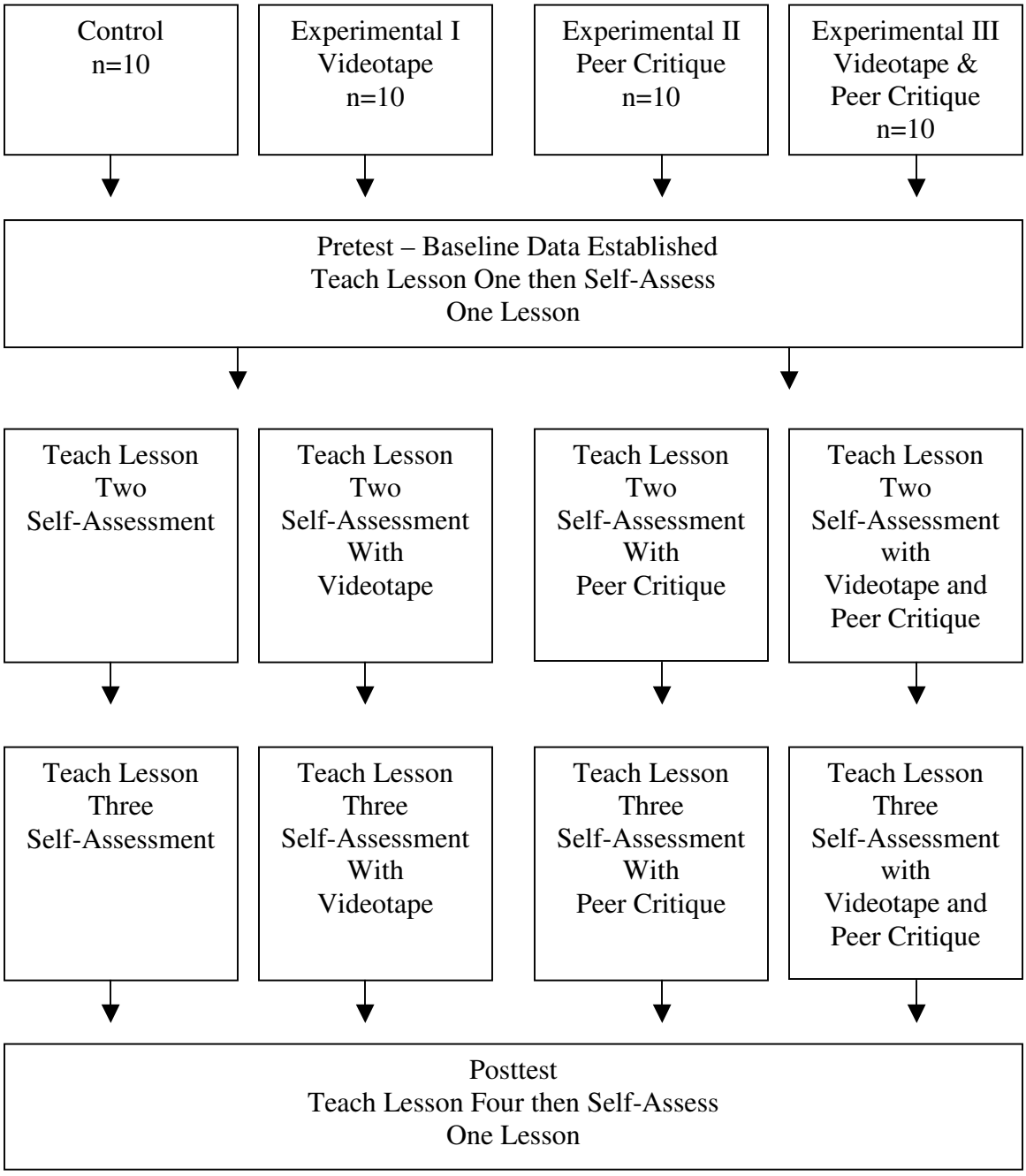
Yarbrough, C., Wapnick, J., & Kelly, R. (1979). Effect of videotape feedback techniques on performance, verbalization, and attitude of beginning conductors. *Journal of Research in Music Education, 27*(2), 103-112.

Zepeda, S. J. (2007). *Instructional supervision: applying tools and concepts* (2nd ed.). Larchmont, NY: Eye On Education.

APPENDICES

APPENDIX A

Study Design



APPENDIX B

Internet Personality Inventory Survey

On the following pages, there are phrases describing people's behaviors. Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex, and roughly your same age. So that you can describe yourself in an honest manner, your responses are anonymous. Please read each statement carefully, and then click the bubble that corresponds to the number on the scale.

1.	Make friends easily.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
2.	Am indifferent to the feelings of others.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
3.	Am exacting in my work.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
4.	Have frequent mood swings.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
5.	Spend time reflecting on things.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
6.	Respect authority.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
7.	Find it difficult to approach others.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
8.	Make people feel at ease.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
9.	Waste my time.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
10.	Get irritated easily.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate

11.	Avoid imposing my will on others.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
12.	Let myself be directed by others.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
13.	Take charge.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
14.	Inquire about others' well-being.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
15.	Do things according to a plan.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
16.	Often feel blue.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
17.	Am full of ideas.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
18.	Do not work as hard as the majority of people around me.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
19.	Don't talk a lot.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
20.	Know how to comfort others.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
21.	Do things in a half-way manner.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
22.	Get angry easily.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐
23.	Will not probe deeply into a subject.	Very Inaccurate ☐	Moderately Inaccurate ☐	Slightly Inaccurate ☐	Slightly Accurate ☐	Moderately Accurate ☐	Very Accurate ☐

24.	Demand to be the center of interest.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
25.	Know how to captivate people.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
26.	Love children.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
27.	Continue until everything is perfect.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
28.	Panic easily.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
29.	Carry the conversation to a higher level.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
30.	Sometimes it is too much of a bother to do exactly what is promised.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
31.	Am the life of the party.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
32.	Insult people.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
33.	Am always prepared.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
34.	Get stressed out easily.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
35.	Have a rich vocabulary.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
36.	Would rather get a bad grade than copy someone else's homework and turn it in as my own.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate

37.	Often feel uncomfortable around others.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
38.	Am interested in people.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
39.	Leave my belongings around.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
40.	Am relaxed most of the time.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
41.	Have difficulty understanding abstract ideas.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
42.	Need a creative outlet.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
43.	Feel comfortable around people.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
44.	Am not interested in other people's problems.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
45.	Pay attention to details.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
46.	Worry about things.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
47.	Have a vivid imagination.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
48.	Demand obedience.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
49.	Keep in the background.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
50.	Sympathize with others' feelings.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate

51.	Make a mess of things.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
52.	Seldom feel blue.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
53.	Am not interested in abstract ideas.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
54.	Like to amuse others.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
55.	Start conversations.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
56.	Feel little concern for others.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
57.	Get chores done right away.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
58.	Am easily disturbed.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
59.	Have excellent ideas.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate
60.	Feel crushed by setbacks.	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate

APPENDIX C

Timetable of Experiment

Week	Activity	Assignment
Fall Semester 2008		
Nov/Dec.	Pilot Study / Class Selection	Select Participants
Spring Semester 2009		
1	Basic Information Session I	Students Consider Participation
2	Basic Information Session II Consent Forms Given and Signed	Participants Gathered & Placed in Group
3	Pretest – ½ Experimental & ½ Control	Complete Observation Form
4	Pretest – ½ Experimental & ½ Control	Complete Observation Form
5	Teaching Episode 1 - ½ Experimental & ½ Control	Complete Observation Form
6	Teaching Episode 1 - ½ Experimental & ½ Control	Complete Observation Form
7	Teaching Episode 2 - ½ Experimental & ½ Control	Complete Observation Form
8	Teaching Episode 2 - ½ Experimental & ½ Control	Complete Observation Form
9	Posttest – ½ Experimental & ½ Control	Complete Observation Form
10	Posttest – ½ Experimental & ½ Control	Complete Observation Form
11	Input Data	N/A
12	Analyze Data	N/A
13	Results / Discussion	N/A
14	Results / Discussion	N/A

APPENDIX D

Evaluation Form One

1. Appropriateness of starting pitch
Too high OK Too Low

2. Matched starting pitch when singing
Yes Sometimes No

3. Used starting pitch consistently
Yes No

4. Sang with accurate pitch throughout song
E G O F P

5. Gave clear instructions
E G O F P

6. Attention to words, pronunciation
E G O F P

7. Gave clear cues for group to start or stop
E G O F P

8. Maintained steady tempo (within each segment/procedure)
E G O F P

9. Tone quality
E G O F P

10. Size of teaching steps (amount of material within one teaching procedure)
E G O F P

11. Logical teaching steps
E G O F P

12. Spoke clearly and understandably
E G O F P

13. Expressive speaking voice
E G O F P

14. Pleasant facial expression
E G O F P

15. Eye contact
E G O F P

16. Enthusiasm/leadership
E G O F P

17. Appropriateness of song choice
E G O F P

18. Students achieved independence
E G O F P

19. appeared well prepared
E G O F P

20. Percentage of approvals
0-35% 35-65% 65-100%

21. Percentage of disapprovals
0-35% 35-65% 65-100%

APPENDIX E

Evaluation Form Two

1. Starting pitch played correctly:
   Yes Sometimes No

2. Matched starting pitch when singing
   Yes Sometimes No

3. Started group correctly (first time)
   Yes No

4. Eye contact when starting group
   Yes No

5. Played without starting over
   Yes No

6. Accurate chords
   Yes No
   If no, # of errors_____

7. Maintained steady beat
   Yes No

8. Played with no pauses between chords
   Yes Mostly No

9. If made mistakes, continued smoothly
   Yes Sometimes No N/A

10. If made mistakes, did not show with face or body
    Yes Sometimes No N/A

11. Played loudly enough to be hear
    Yes Almost No

12. Pleasant facial expression
    Excellent Good OK Fair Poor

13. Enthusiasm
    Excellent Good OK Fair Poor

14. Leadership
    Excellent Good OK Fair Poor

15. Eye contact throughout song
    Excellent Good OK Fair Poor

16. Appeared well-prepared
    Excellent Good OK Fair Poor

APPENDIX F

Evaluation Form Three

Your Name:_____ Person Teaching:_____

Teaching

1. Clarity of Instruction

Excellent Satisfactory Needs Impr. N/A

2. Organization

Excellent Satisfactory Needs Impr. N/A

3. Logical Sequencing

Excellent Satisfactory Needs Impr. N/A

4. Size Teaching Steps

Excellent Satisfactory Needs Impr. N/A

5. Pacing

Excellent Satisfactory Needs Impr. N/A

6. Accuracy of Concept

Excellent Satisfactory Needs Impr. N/A

7. Student Involvement

Excellent Satisfactory Needs Impr. N/A

8. Methods & Activities Appropriate

Excellent Satisfactory Needs Impr. N/A

9. Visual Aids

Excellent Satisfactory Needs Impr. N/A

10. Students Achieve Independence

Excellent Satisfactory Needs Impr. N/A

11. Preparation

Excellent Satisfactory Needs Impr. N/A

12. Creativity

Excellent Satisfactory Needs Impr. N/A

13. Eye Contact while Singing

Excellent Satisfactory Needs Impr. N/A

14. Eye Contact while Playing

Excellent Satisfactory Needs Impr. N/A

15. Eye Contact while Speaking

Personality

Excellent Satisfactory Needs Impr. N/A

16. Clarity of Speech

Excellent Satisfactory Needs Impr. N/A

17. Expressive Voice

Excellent Satisfactory Needs Impr. N/A

18. Expressive Face

Excellent Satisfactory Needs Impr. N/A

19. Enthusiasm

Excellent Satisfactory Needs Impr. N/A

20. Leadership

Excellent Satisfactory Needs Impr. N/A

21. Appropriate feedback in Class

Excellent Satisfactory Needs Impr. N/A

22. Positive Verbal Interaction

Excellent Satisfactory Needs Impr. N/A

Music Skills

23. Matched Maintained Starting Pitch

Excellent Satisfactory Needs Impr. N/A

24. Maintained Pitch Accuracy

Excellent Satisfactory Needs Impr. N/A

25. Continued Smoothly after Mistake

Excellent Satisfactory Needs Impr. N/A

26. Pitch Level of Song

Excellent Satisfactory Needs Impr. N/A

27. Clarity of Group Start or Stop Cues

Excellent Satisfactory Needs Impr. N/A

28. Accompaniment Effect

Excellent Satisfactory Needs Impr. N/A

29. Accurate Rhythm/Beat

Excellent Satisfactory Needs Impr. N/A

APPENDIX G

Evaluation Form Four

Your Name:_____ Person Teaching:_____

Teaching

1. Clarity of Instruction
Excellent Satisfactory Needs Impr. N/A
2. Organization
Excellent Satisfactory Needs Impr. N/A
3. Logical Sequencing
Excellent Satisfactory Needs Impr. N/A
4. Size Teaching Steps
Excellent Satisfactory Needs Impr. N/A
5. Pacing
Excellent Satisfactory Needs Impr. N/A
6. Accuracy of Concept
Excellent Satisfactory Needs Impr. N/A
7. Student Involvement
Excellent Satisfactory Needs Impr. N/A
8. Methods & Activities Appropriate
Excellent Satisfactory Needs Impr. N/A
9. Visual Aids
Excellent Satisfactory Needs Impr. N/A
10. Students Achieve Independence
Excellent Satisfactory Needs Impr. N/A
11. Preparation
Excellent Satisfactory Needs Impr. N/A
12. Creativity
Excellent Satisfactory Needs Impr. N/A
13. Eye Contact while Singing
Excellent Satisfactory Needs Impr. N/A
14. Eye Contact while Playing
Excellent Satisfactory Needs Impr. N/A
15. Eye Contact while Speaking
Excellent Satisfactory Needs Impr. N/A

Personality

16. Clarity of Speech
Excellent Satisfactory Needs Impr. N/A
17. Expressive Voice
Excellent Satisfactory Needs Impr. N/A
18. Expressive Face
Excellent Satisfactory Needs Impr. N/A
19. Enthusiasm
Excellent Satisfactory Needs Impr. N/A
20. Leadership
Excellent Satisfactory Needs Impr. N/A
21. Appropriate feedback in Class
Excellent Satisfactory Needs Impr. N/A
22. Positive Verbal Interaction
Excellent Satisfactory Needs Impr. N/A

Music Skills

23. Matched Maintained Starting Pitch
Excellent Satisfactory Needs Impr. N/A
24. Maintained Pitch Accuracy
Excellent Satisfactory Needs Impr. N/A
25. Continued Smoothly after Mistake
Excellent Satisfactory Needs Impr. N/A
26. Pitch Level of Song
Excellent Satisfactory Needs Impr. N/A
27. Clarity of Group Start or Stop Cues
Excellent Satisfactory Needs Impr. N/A
28. Accompaniment Effect
Excellent Satisfactory Needs Impr. N/A
29. Accurate Rhythm/Beat
Excellent Satisfactory Needs Impr. N/A

APPENDIX H

Personality Profile

What follows is the results of your survey responses. The results here are grouped into five categories: extraversion, agreeableness, conscientiousness, emotional stability, and openness. These categories represent the way that most people talk about personality and so they may reflect cultural or social biases.

While many or all of these categories may look like words you typically use (even ones that often are accompanied with a value judgment) it is important to understand that these five factors are really labels used by psychologists to describe differences between people.

This is not a psycho-analysis; the results presented here were created directly from your responses to the items. For that reason, it is unlikely that there should be a miss-match between our descriptions and how you or others view themselves.

Extraversion	Low		High
Agreeableness	Low		High
Conscientiousness	Low		High
Emotional Stability	Low		High
Openness	Low		High
	Percentile	0 --------- 25 --------- 50 --------- 75 -------- 100	

Extraversion Report

Extraversion is marked by pronounced engagement with the external world. Extraverts enjoy being with people, are full of energy, and often experience positive emotions. They tend to be enthusiastic, action-oriented, individuals who are likely to say "Yes!" or "Let's go!" to opportunities for excitement. In groups they like to talk, assert themselves, and draw attention to themselves.

Introverts lack the exuberance, energy, and activity levels of extraverts. They tend to be quiet, low-key, deliberate, and disengaged from the social world. Their lack of social involvement should not be interpreted as shyness or depression; the introvert simply needs less stimulation than an extravert and prefers to be alone. The independence and reserve of the introvert is sometimes mistaken as unfriendliness or arrogance. In reality, an introvert who scores high on the agreeableness dimension will not seek others out but will be quite pleasant when approached.

Score at a Glance:	Total Score: 38	Average Response: 3.6

Your average score on extraversion was 3.6, which is considered average. It is in approximately the 38th percentile for males under the age of 21.

Your score on Extraversion is average, indicating you are neither a subdued loner nor a jovial chatterbox. You enjoy time with others but also time alone.

Agreeableness Report

Agreeableness reflects individual differences in concern with cooperation and social harmony. Agreeable individuals value getting along with others. They are therefore considerate, friendly, generous, helpful, and willing to compromise their interests with others'. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy.

Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being, and therefore are unlikely to extend themselves for other people. Sometimes their skepticism about others' motives causes them to be suspicious, unfriendly, and uncooperative.

Agreeableness is obviously advantageous for attaining and maintaining popularity. Agreeable people are better liked than disagreeable people. On the other hand, agreeableness is not useful in situations that require tough or absolute objective decisions. Disagreeable people can make excellent scientists, critics, or soldiers.

Score at a Glance:	Total Score: 66	Average Response: 4.2

Your average score on agreeableness was 4.2, which is considered average. It is in approximately the 66th percentile for males under the age of 21.

Your level of Agreeableness is average, indicating some concern with others' Needs, but, generally, unwillingness to sacrifice yourself for others.

Conscientiousness Report

Conscientiousness concerns the way in which we control, regulate, and direct our impulses. Impulses are not inherently bad; occasionally time constraints require a snap decision, and acting on our first impulse can be an effective response. Also, in times of play rather than work, acting spontaneously and impulsively can be fun. Others can see impulsive individuals as colorful, fun-to-be-with, and zany.

Nonetheless, acting on impulse can lead to trouble in a number of ways. Some impulses are antisocial. Uncontrolled antisocial acts not only harm other members of society, but also can result in retribution toward the perpetrator of such impulsive acts. Another problem with impulsive acts is that they often produce immediate rewards but undesirable, long-term consequences. Examples include excessive socializing that leads to being fired from one's job, hurling an insult that causes the breakup of an important relationship, or using pleasure-inducing drugs that eventually destroy one's health.

Impulsive behavior, even when not seriously destructive, diminishes a person's effectiveness in significant ways. Acting impulsively disallows contemplating alternative courses of action, some of which would have been wiser than the impulsive choice. Impulsivity also sidetracks people during projects that require organized sequences of steps or stages. Accomplishments of an impulsive person are therefore small, scattered, and inconsistent.

A hallmark of intelligence, what potentially separates human beings from earlier life forms, is the ability to think about future consequences before acting on an impulse. Intelligent activity involves contemplation of long-range goals, organizing and planning routes to these goals, and persisting toward one's goals in the face of short-lived impulses to the contrary. The idea that intelligence involves impulse control is nicely captured by the term prudence, an alternative label for the Conscientiousness domain. Prudent means both wise and cautious. Persons who score high on the Conscientiousness scale are, in fact, perceived by others as intelligent.

The benefits of high conscientiousness are obvious. Conscientious individuals avoid trouble and achieve high levels of success through purposeful planning and persistence. They are also positively regarded by others as intelligent and reliable. On the negative side, they can be compulsive perfectionists and workaholics. Furthermore, extremely conscientious individuals might be regarded as stuffy and boring. Unconscientious people may be criticized for their unreliability, lack of ambition, and failure to stay within the lines, but they will experience many short-lived pleasures and they will never be called stuffy.

Score at a Glance:	Total Score: 99	Average Response: 6

Your average score on conscientiousness was 6, which is considered high. It is in approximately the 99th percentile for males under the age of 21.

Your score on Conscientiousness is high. This means you set clear goals and pursue them with determination. People regard you as reliable and hard-working.

Emotional Stability Report

Emotional stability is the opposite of emotional reactivity, which is the tendency to experience negative feelings. Those who score low on emotional stability may experience primarily one specific negative feeling such as anxiety, anger, or depression, but are likely to experience several of these emotions. People low in emotional stability are emotionally reactive. They respond emotionally to events that would not affect most people, and their reactions tend to be more intense than normal. They are more likely to interpret ordinary situations as threatening, and minor frustrations as hopelessly difficult. Their negative emotional reactions tend to persist for unusually long periods of time, which means they are often in a bad mood. These problems in emotional regulation can diminish a ones ability to think clearly, make decisions, and cope effectively with stress.

At the other end of the scale, individuals who score high in emotional stability are less easily upset and are less emotionally reactive. They tend to be calm, emotionally stable,

and free from persistent negative feelings. Freedom from negative feelings does not mean that high scorers experience a lot of positive feelings; frequency of positive emotions is a component of the Extraversion domain.

Score at a Glance:	Total Score: 62	Average Response: 4.1

Your average score on emotional stability was 4.1, which is considered average. It is in approximately the 62nd percentile for males under the age of 21.

Your score on Emotional Stability is average, indicating that your level of emotional reactivity is typical of the general population. Stressful and frustrating situations are somewhat upsetting to you, but you are generally able to get over these feelings and cope with these situations.

Openness Report

Openness to Experience describes a dimension of cognitive style that distinguishes imaginative, creative people from down-to-earth, conventional people. Open people are intellectually curious, appreciative of art, and sensitive to beauty. They tend to be, compared to closed people, more aware of their feelings. They tend to think and act in individualistic and nonconforming ways. Intellectuals typically score high on Openness to Experience; consequently, this factor has also been called Culture or Intellect. Nonetheless, Intellect is probably best regarded as one aspect of openness to experience. Scores on Openness to Experience are only modestly related to years of education and scores on standard intelligent tests.

Another characteristic of the open cognitive style is a facility for thinking in symbols and abstractions far removed from concrete experience. Depending on the individual's specific intellectual abilities, this symbolic cognition may take the form of mathematical, logical, or geometric thinking, artistic and metaphorical use of language, music composition or performance, or one of the many visual or performing arts. People with low scores on openness to experience tend to have narrow, common interests. They prefer the plain, straightforward, and obvious over the complex, ambiguous, and subtle. They may regard the arts and sciences with suspicion, regarding these endeavors as abstruse or of no practical use. Closed people prefer familiarity over novelty; they are conservative and resistant to change.

Openness is often presented as healthier or more mature by psychologists, who are often themselves open to experience. However, open and closed styles of thinking are useful in different environments. The intellectual style of the open person may serve a professor well, but research has shown that closed thinking is related to superior job performance in police work, sales, and a number of service occupations.

Score at a Glance:	Total Score: 66	Average Response: 4.5

Your average score on openness was 4.5, which is considered average. It is in approximately the 66th percentile for males under the age of 21. Your score on Openness to Experience is average, indicating you enjoy tradition but are willing to try new things. Your thinking is neither simple nor complex. To others you appear to be a well-educated person but not an intellectual.

APPENDIX I

Teaching Lesson One (Pretest): Full Data

Participant	Group	Possible	SE	SE%	PE	PE%	PE% - SE% (*d*)
3	C	88	75	85.23	67	76.14	9.09
4	C	88	57	64.77	68	77.27	12.5
5	C	88	82	93.18	68	77.27	15.91
11	C	88	76	86.36	71	80.68	5.5
12	C	88	77	87.5	71	80.68	6.82
16	C	88	75	85.23	72	81.82	3.41
23	C	88	81	92.04	80	90.91	1.13
29	C	88	70	79.55	85	96.59	17.04
30	C	88	71	80.68	85	96.59	15.91
33	C	88	82	93.18	86	97.73	4.55
7	E1	88	70	79.55	69	78.41	1.14
8	E1	88	65	73.86	71	80.68	6.82
9	E1	88	69	78.41	71	80.68	2.27
15	E1	88	71	80.68	72	81.82	1.14
18	E1	88	78	88.64	72	81.82	6.82
20	E1	88	77	87.5	79	89.77	2.27
22	E1	88	79	89.77	80	90.91	1.14
25	E1	88	72	81.82	81	92.04	10.22
35	E1	88	73	82.95	87	98.86	15.91
37	E1	88	80	90.91	87	98.86	7.95
1	E2	88	79	89.77	64	72.73	17.04
6	E2	88	69	78.41	69	78.41	0
17	E2	88	77	87.5	72	81.82	5.68
19	E2	88	65	73.86	76	86.36	12.5
24	E2	88	64	72.73	81	92.04	19.31
27	E2	88	80	90.91	84	95.45	4.54
31	E2	88	80	90.91	85	96.59	5.68
32	E2	88	83	94.32	85	96.59	2.27
36	E2	88	73	82.95	87	98.86	15.91
38	E2	88	67	76.14	88	100	23.86
2	E3	88	70	79.55	66	75	4.55
10	E3	88	71	80.68	71	80.68	0
13	E3	88	80	90.91	71	80.68	10.23
14	E3	88	66	75	72	81.82	6.82
21	E3	88	83	94.32	79	89.77	4.55
26	E3	88	79	89.77	83	94.32	4.55
28	E3	88	84	95.45	84	95.45	0
34	E3	88	71	80.68	87	98.86	18.18
39	E3	88	79	89.77	88	100	10.23
40	E3	88	81	92.04	88	100	7.96

APPENDIX J

Teaching Lesson Two: Full Data

Participant	Group	Possible	SE	SE%	PE	PE%	PE% - SE% (*d*)
3	C	55	46	83.64	48	87.27	3.63
4	C	55	42	76.36	47	85.45	9.09
5	C	55	43	78.18	48	87.27	9.09
11	C	55	52	94.56	47	85.45	9.11
12	C	55	48	87.27	48	87.27	0
16	C	55	47	85.45	48	87.27	1.82
23	C	55	47	85.45	33	60	25.45
29	C	55	47	85.45	43	78.18	7.27
30	C	55	44	80	43	78.18	1.82
33	C	55	52	94.56	49	89.09	5.47
7	E1	55	47	85.45	44	80	5.45
8	E1	55	51	92.73	48	87.27	5.46
9	E1	55	51	92.73	45	81.82	10.91
15	E1	55	52	94.56	48	87.27	7.29
18	E1	55	50	90.91	45	81.82	9.09
20	E1	55	47	85.45	50	90.91	5.46
22	E1	55	51	92.73	43	78.18	14.55
25	E1	55	51	92.73	53	96.36	3.63
35	E1	55	48	87.27	48	87.27	0
37	E1	55	55	100	49	89.09	10.91
1	E2	55	48	87.27	39	70.91	16.36
6	E2	55	46	83.64	48	87.27	3.63
17	E2	55	49	89.09	48	87.27	1.82
19	E2	55	41	74.55	48	87.27	12.72
24	E2	55	41	74.55	49	89.09	14.54
27	E2	55	46	83.64	39	70.91	12.73
31	E2	55	47	85.45	46	83.64	1.81
32	E2	55	52	94.56	52	94.56	0
36	E2	55	49	89.09	53	96.36	7.27
38	E2	55	49	89.09	48	87.27	1.82
2	E3	55	45	81.82	46	83.64	1.82
10	E3	55	45	81.82	50	90.91	9.09
13	E3	55	44	80	48	87.27	7.27
14	E3	55	41	74.55	48	87.27	12.73
21	E3	55	50	90.91	43	78.18	12.72
26	E3	55	49	89.09	49	89.09	0
28	E3	55	55	100	46	83.64	16.36
34	E3	55	38	69.1	46	83.64	14.54
39	E3	55	52	94.56	52	94.56	0
40	E3	55	47	85.45	52	94.56	9.11

APPENDIX K

Teaching Lesson Three: Full Data

Participant	Group	Possible	SE	SE%	PE	PE%	PE% - SE% (*d*)
3	C	156	101	64.74	80	51.28	13.46
4	C	156	86	55.13	79	50.64	4.49
5	C	156	86	55.13	77	49.36	5.77
11	C	156	82	52.56	71	45.51	7.05
12	C	156	79	50.64	79	50.64	0
16	C	156	93	59.62	80	51.28	8.34
23	C	156	105	67.31	86	55.13	12.18
29	C	156	85	54.49	76	48.72	5.77
30	C	156	100	64.1	85	54.49	9.61
33	C	156	105	67.31	78	50	17.31
7	E1	156	88	56.41	81	51.92	4.49
8	E1	156	97	62.18	81	51.92	10.26
9	E1	156	99	63.46	82	52.56	10.9
15	E1	156	84	53.85	70	50	3.85
18	E1	156	103	66.03	81	51.92	14.11
20	E1	156	107	68.59	82	52.56	16.03
22	E1	156	79	50.64	51	32.69	17.95
25	E1	156	111	71.15	80	51.28	19.87
35	E1	156	96	61.54	82	52.56	8.98
37	E1	156	110	70.51	84	53.85	16.66
1	E2	156	97	62.18	79	50.64	11.54
6	E2	156	94	60.26	79	50.64	9.62
17	E2	156	96	61.54	80	51.28	10.26
19	E2	156	80	51.28	72	46.15	5.13
24	E2	156	96	61.54	78	50	11.54
27	E2	156	88	56.41	70	44.87	11.54
31	E2	156	110	70.51	81	51.92	18.59
32	E2	156	109	69.87	78	50	19.87
36	E2	156	102	65.38	83	53.21	12.17
38	E2	156	83	53.21	78	50	3.21
2	E3	156	101	64.74	82	52.56	12.18
10	E3	156	84	53.85	79	50.64	3.21
13	E3	156	82	52.56	68	43.59	8.97
14	E3	156	90	57.69	83	53.21	4.48
21	E3	156	98	62.82	83	53.21	9.61
26	E3	156	110	70.51	81	51.92	18.59
28	E3	156	115	73.72	79	50.64	23.08
34	E3	156	108	69.23	84	53.85	15.38
39	E3	156	108	69.23	82	52.56	16.67
40	E3	156	97	62.18	83	53.21	8.97

APPENDIX L

Teaching Lesson Four (Posttest): Full Data

Participant	Group	Possible	SE	SE%	PE	PE%	PE% - SE% (*d*)
3	C	156	100	64.1	81	51.92	12.18
4	C	156	96	61.54	79	50.64	10.9
5	C	156	93	59.62	78	50	9.62
11	C	156	104	66.67	82	52.56	14.11
12	C	156	81	51.92	85	54.49	-2.57
16	C	156	90	57.69	79	50.64	7.05
23	C	156	105	67.31	86	55.13	12.18
29	C	156	90	57.69	65	41.66	16.03
30	C	156	103	66.03	83	53.21	12.82
33	C	156	105	67.31	78	50	17.31
7	E1	156	90	57.69	75	48.08	9.61
8	E1	156	95	60.9	76	48.72	12.18
9	E1	156	83	53.21	82	52.56	0.65
15	E1	156	91	58.33	79	50.64	7.69
18	E1	156	82	52.56	82	52.56	0
20	E1	156	107	68.59	82	52.56	16.03
22	E1	156	79	50.64	51	32.69	17.95
25	E1	156	109	69.87	80	51.28	18.59
35	E1	156	90	57.69	81	51.92	5.77
37	E1	156	110	70.51	84	53.85	16.66
1	E2	156	92	58.97	81	51.92	7.05
6	E2	156	96	61.54	62	39.74	21.8
17	E2	156	83	53.21	81	51.92	1.29
19	E2	156	80	51.28	77	49.36	1.92
24	E2	156	96	61.54	78	50	11.54
27	E2	156	109	69.87	77	49.36	20.51
31	E2	156	98	62.82	79	50.64	12.18
32	E2	156	106	67.95	78	50	17.95
36	E2	156	97	62.18	76	48.72	13.46
38	E2	156	90	57.69	78	50	7.69
2	E3	156	104	66.67	83	53.21	13.46
10	E3	156	78	50	64	41.03	8.97
13	E3	156	86	55.13	77	49.36	5.77
14	E3	156	96	61.54	79	50.64	10.9
21	E3	156	98	62.82	83	53.21	9.61
26	E3	156	107	68.59	82	52.56	16.03
28	E3	156	113	72.44	79	50.64	21.8
34	E3	156	103	66.03	80	51.28	14.75
39	E3	156	103	66.03	82	52.56	13.47
40	E3	156	112	71.79	83	53.21	18.58

APPENDIX M

Individual Scores: Control Group

Lesson One		
Self Evaluation	Professional Evaluation	Difference (*d*)
92.04	90.91	1.13
85.23	81.82	3.41
93.18	97.73	4.55
86.36	80.68	5.5
87.5	80.68	6.82
85.23	76.14	9.09
64.77	77.27	12.5
93.18	77.27	15.91
80.68	96.59	15.91
79.55	96.59	17.04

Lesson Two		
Self Evaluation	Professional Evaluation	Difference (*d*)
87.27	87.27	0
85.45	87.27	1.82
80	78.18	1.82
83.64	87.27	3.63
94.56	89.09	5.47
85.45	78.18	7.27
76.36	85.45	9.09
78.18	87.27	9.09
94.56	85.45	9.11
85.45	60	25.45

Lesson Three		
Self Evaluation	Professional Evaluation	Difference (*d*)
50.64	50.64	0
55.13	50.64	4.49
55.13	49.36	5.77
54.49	48.72	5.77
52.56	45.51	7.05
59.62	51.28	8.34
64.1	54.49	9.61
67.31	55.13	12.18
64.74	51.28	13.46
67.31	50	17.31

Lesson Four		
Self Evaluation	Professional Evaluation	Difference (*d*)
51.92	54.49	2.57
57.69	50.64	7.05
59.62	50	9.62
61.54	50.64	10.9
64.1	51.92	12.18
67.31	55.13	12.18
66.03	53.21	12.82
66.67	52.56	14.11
57.69	41.66	16.03
67.31	50	17.31

APPENDIX N

Individual Scores: Experimental Group I

Lesson One		
Self Evaluation	Professional Evaluation	Difference (*d*)
79.55	78.41	1.14
80.68	81.82	1.14
89.77	90.91	1.14
78.41	80.68	2.27
87.5	89.77	2.27
73.86	80.68	6.82
88.64	81.82	6.82
90.91	98.86	7.95
81.82	92.04	10.22
82.95	98.86	15.91

Lesson Two		
Self Evaluation	Professional Evaluation	Difference (*d*)
87.27	87.27	0
92.73	96.36	3.63
85.45	80	5.45
92.73	87.27	5.46
85.45	90.91	5.46
94.56	87.27	7.29
90.91	81.82	9.09
92.73	81.82	10.91
100	89.09	10.91
92.73	78.18	14.55

Lesson Three		
Self Evaluation	Professional Evaluation	Difference (*d*)
53.85	50	3.85
56.41	51.92	4.49
61.54	52.56	8.98
62.18	51.92	10.26
63.46	52.56	10.9
66.03	51.92	14.11
68.59	52.56	16.03
70.51	53.85	16.66
50.64	32.69	17.95
71.15	51.28	19.87

Lesson Four		
Self Evaluation	Professional Evaluation	Difference (*d*)
52.56	52.56	0
53.21	52.56	0.65
57.69	51.92	5.77
58.33	50.64	7.69
57.69	48.08	9.61
60.9	48.72	12.18
68.59	52.56	16.03
70.51	53.85	16.66
50.64	32.69	17.95
69.87	51.28	18.59

APPENDIX O

Individual Scores: Experimental Group II

Lesson One		
Self Evaluation	Professional Evaluation	Difference (*d*)
78.41	78.41	0
94.32	96.59	2.27
90.91	95.45	4.54
87.5	81.82	5.68
90.91	96.59	5.68
73.86	86.36	12.5
82.95	98.86	15.91
89.77	72.73	17.04
72.73	92.04	19.31
76.14	100	23.86

Lesson Two		
Self Evaluation	Professional Evaluation	Difference (*d*)
94.56	94.56	0
85.45	83.64	1.81
89.09	87.27	1.82
89.09	87.27	1.82
83.64	87.27	3.63
89.09	96.36	7.27
74.55	87.27	12.72
83.64	70.91	12.73
74.55	89.09	14.54
87.27	70.91	16.36

Lesson Three		
Self Evaluation	Professional Evaluation	Difference (*d*)
53.21	50	3.21
51.28	46.15	5.13
60.26	50.64	9.62
61.54	51.28	10.26
62.18	50.64	11.54
61.54	50	11.54
56.41	44.87	11.54
65.38	53.21	12.17
70.51	51.92	18.59
69.87	50	19.87

Lesson Four		
Self Evaluation	Professional Evaluation	Difference (*d*)
53.21	51.92	1.29
51.28	49.36	1.92
58.97	51.92	7.05
57.69	50	7.69
61.54	50	11.54
62.82	50.64	12.18
62.18	48.72	13.46
67.95	50	17.95
69.87	49.36	20.51
61.54	39.74	21.8

APPENDIX P

Individual Scores: Experimental Group III

Lesson One		
Self Evaluation	Professional Evaluation	Difference (*d*)
80.68	80.68	0
95.45	95.45	0
79.55	75	4.55
94.32	89.77	4.55
89.77	94.32	4.55
75	81.82	6.82
92.04	100	7.96
90.91	80.68	10.23
89.77	100	10.23
80.68	98.86	18.18

Lesson Two		
Self Evaluation	Professional Evaluation	Difference (*d*)
89.09	89.09	0
94.56	94.56	0
81.82	83.64	1.82
80	87.27	7.27
81.82	90.91	9.09
85.45	94.56	9.11
90.91	78.18	12.72
74.55	87.27	12.73
69.1	83.64	14.54
100	83.64	16.36

Lesson Three		
Self Evaluation	Professional Evaluation	Difference (*d*)
53.85	50.64	3.21
57.69	53.21	4.48
52.56	43.59	8.97
62.18	53.21	8.97
62.82	53.21	9.61
64.74	52.56	12.18
69.23	53.85	15.38
69.23	52.56	16.67
70.51	51.92	18.59
73.72	50.64	23.08

Lesson Four		
Self Evaluation	Professional Evaluation	Difference (*d*)
55.13	49.36	5.77
50	41.03	8.97
62.82	53.21	9.61
61.54	50.64	10.9
66.67	53.21	13.46
66.03	52.56	13.47
66.03	51.28	14.75
68.59	52.56	16.03
71.79	53.21	18.58
72.44	50.64	21.8

APPENDIX Q

ANOVA Data Comparing Differences in Groups Self-assessment Scores

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	207.635027	69.211676	0.79	0.5095
Error	36	3169.084550	88.030126		
Corrected Total	39	3376.719577			

R-Square	Coeff Var	Root MSE	diff Mean
0.061490	253.1859	9.382437	3.705750

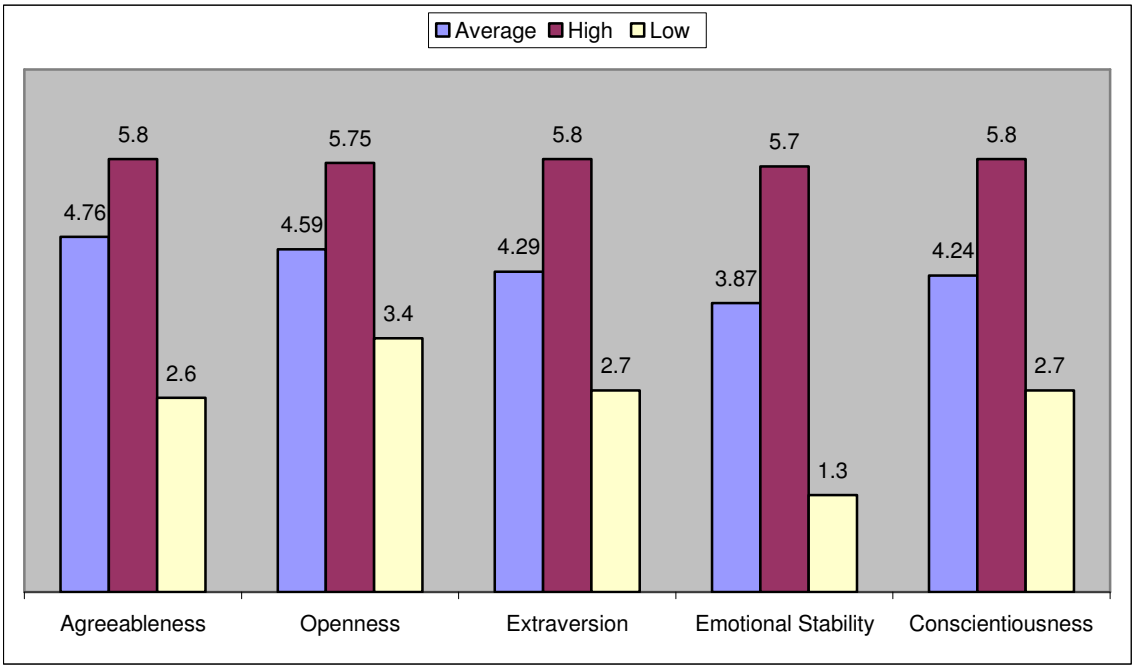
Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	207.6350275	69.2116758	0.79	0.5095

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	88.03013
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	11.301

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**group**
A	6.727	10	E3
A			
A	4.945	10	E1
A			
A	2.291	10	C
A			
A	0.860	10	E2

APPENDIX R

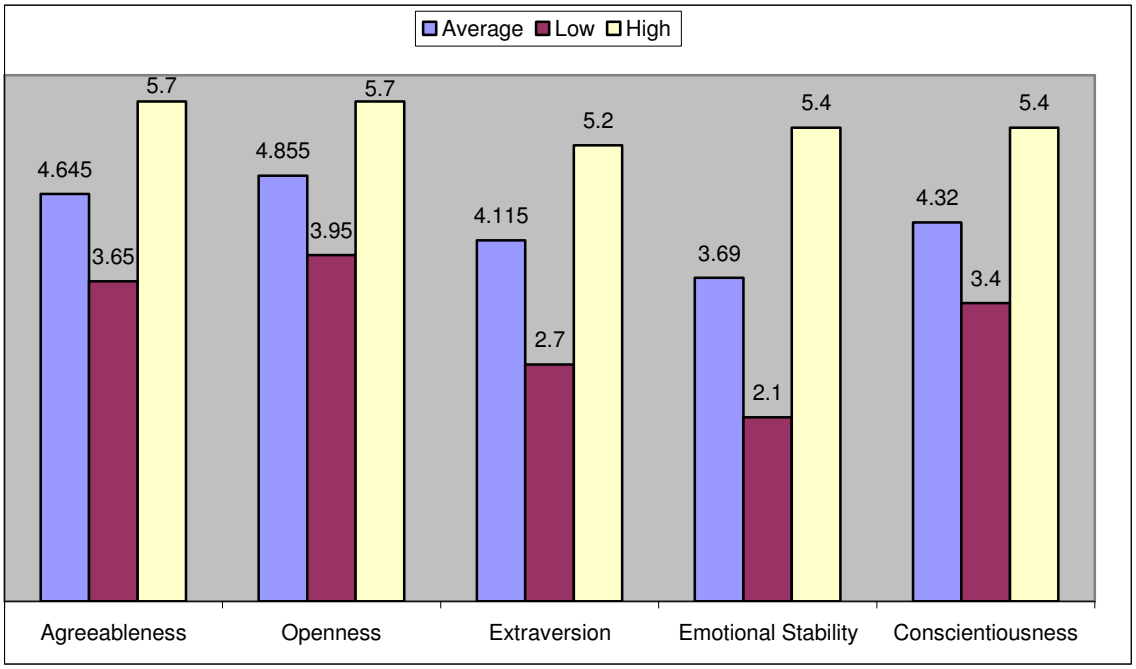
Personality Profile All Groups: Average, High, and Low Scores, and Standard Deviation



Personality Trait		Standard Deviation
Agreeableness		0.6467
Openness		0.6459
Extraversion		0.7744
Emotional Stability		0.8686
Conscientiousness		0.8524

APPENDIX S

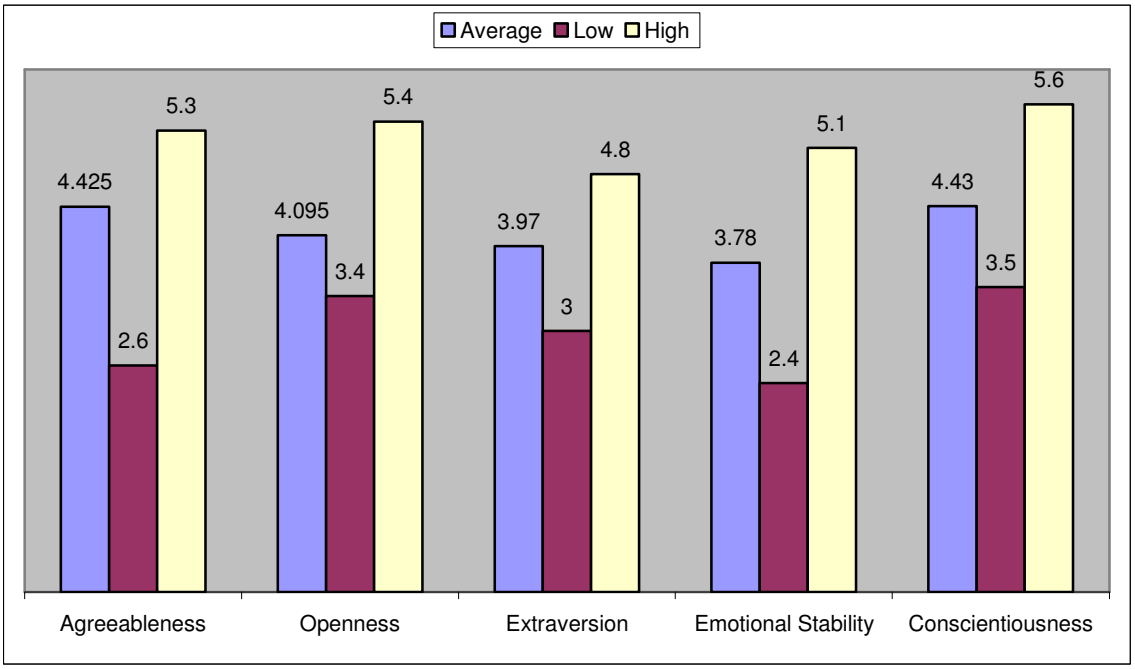
Personality Profile Control Group: Average, High, and Low Scores, and Standard Deviation



Personality Trait		Standard Deviation
Agreeableness		0.6568
Openness		0.5956
Extraversion		0.8570
Emotional Stability		0.9620
Conscientiousness		0.6579

APPENDIX T

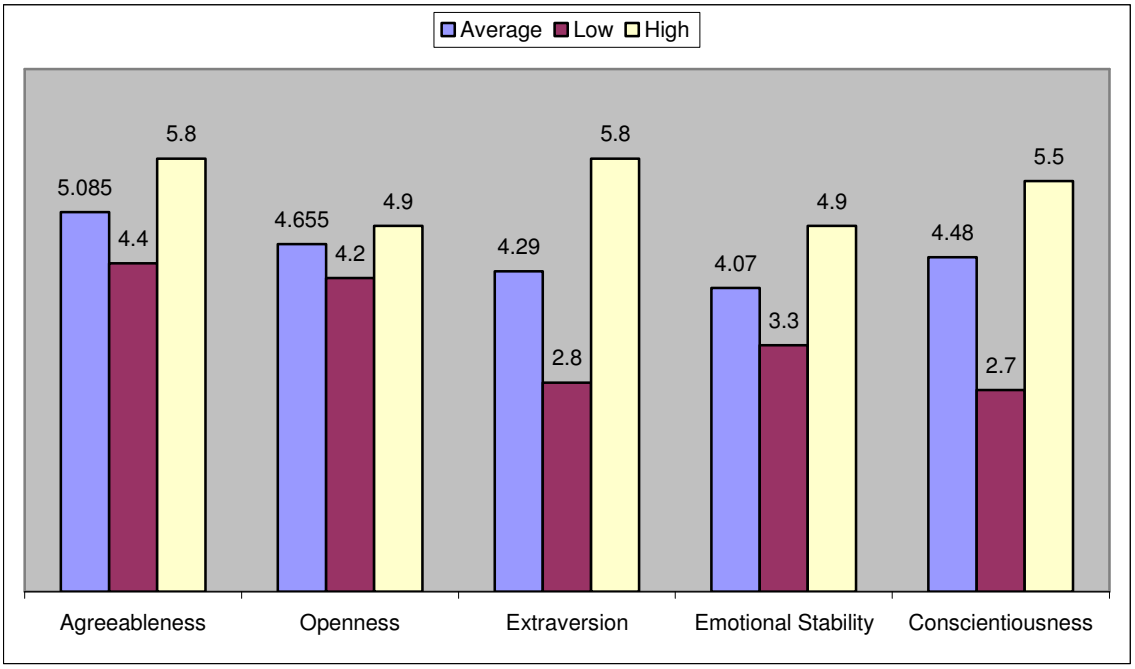
Personality Profile Experimental Group I: Average, High, and Low Scores, and Standard Deviation



Personality Trait		Standard Deviation
Agreeableness		0.8059
Openness		0.7057
Extraversion		0.6516
Emotional Stability		0.7130
Conscientiousness		0.7931

APPENDIX U

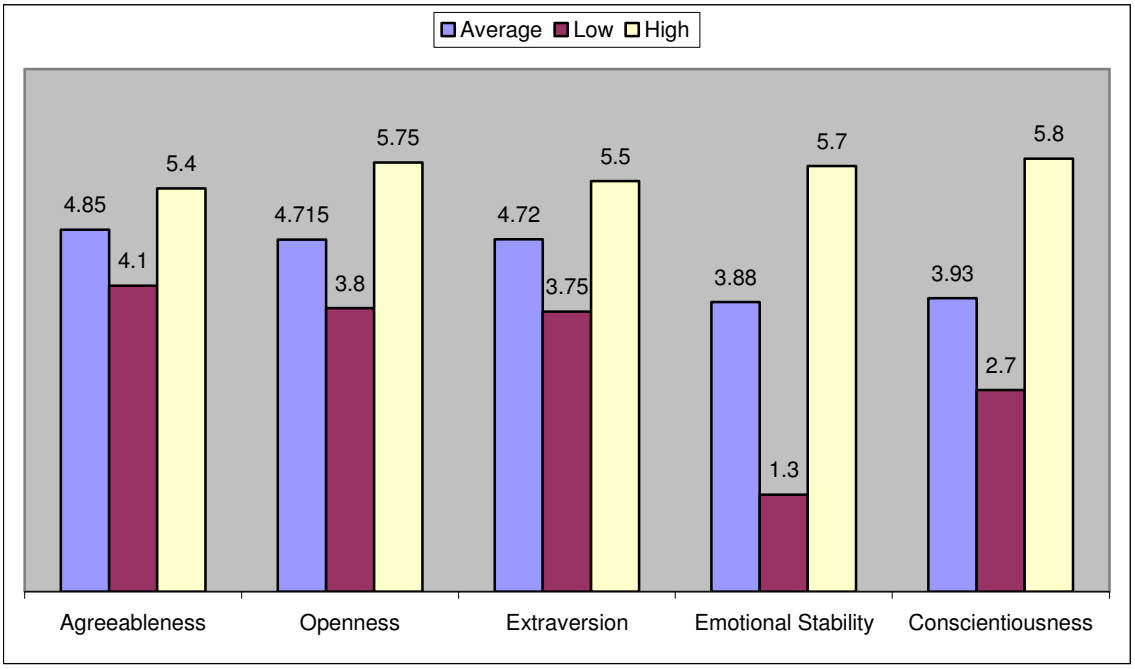
Personality Profile Experimental Group II: Average, High, and Low Scores, and Standard Deviation



Personality Trait		Standard Deviation
Agreeableness		0.5045
Openness		0.2166
Extraversion		0.8962
Emotional Stability		0.5438
Conscientiousness		0.8791

APPENDIX V

Personality Profile Experimental Group III: Average, High, and Low Scores, and Standard Deviation



Personality Trait		Standard Deviation
Agreeableness		0.4601
Openness		0.7326
Extraversion		0.5334
Emotional Stability		1.2047
Conscientiousness		1.052

APPENDIX W

ANOVA Data Comparing Personality Trait: AGREEABLENESS Amongst Groups

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.38868750	0.79622917	2.06	0.1230
Error	36	13.92375000	0.38677083		
Corrected Total	39	16.31243750			

R-Square	Coeff Var	Root MSE	ag Mean
0.146434	13.08938	0.621909	4.751250

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	2.38868750	0.79622917	2.06	0.1230

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	0.386771
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	0.7491

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**group**
A	5.0850	10	E2
A			
A	4.8500	10	E3
A			
A	4.6450	10	C
A			
A	4.4250	10	E1

APPENDIX X

ANOVA Data Comparing Personality Trait: EXTRAVERSION Amongst Groups

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.16868750	1.05622917	1.88	0.1503
Error	36	20.22125000	0.56170139		
Corrected Total	39	23.38993750			

R-Square	Coeff Var	Root MSE	ex Mean
0.135472	17.53653	0.749467	4.273750

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	3.16868750	1.05622917	1.88	0.1503

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	0.561701
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	0.9027

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**group**
A	4.7200	10	E3
A			
A	4.2900	10	E2
A			
A	4.1150	10	C
A			
A	3.9700	10	E1

APPENDIX Y

ANOVA Data Comparing Personality Trait: EMOTIONAL STABILITY Amongst Groups

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.79700000	0.26566667	0.33	0.8007
Error	36	28.62700000	0.79519444		
Corrected Total	39	29.42400000			

R-Square	Coeff Var	Root MSE	em Mean
0.027087	23.13195	0.891737	3.855000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	0.79700000	0.26566667	0.33	0.8007

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	0.795194
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	1.0741

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**group**
A	4.0700	10	E2
A			
A	3.8800	10	E3
A			
A	3.7800	10	E1
A			
A	3.6900	10	C

APPENDIX Z

ANOVA Data Comparing Personality Trait: CONSCIENTIOUSNESS Amongst Groups

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.86200000	0.62066667	0.84	0.4789
Error	36	26.47400000	0.73538889		
Corrected Total	39	28.33600000			

R-Square	Coeff Var	Root MSE	co Mean
0.065711	19.98947	0.857548	4.290000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	1.86200000	0.62066667	0.84	0.4789

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	0.735389
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	1.0329

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**group**
A	4.4800	10	E2
A			
A	4.4300	10	E1
A			
A	4.3200	10	C
A			
A	3.9300	10	E3

APPENDIX AA

ANOVA Data Comparing Personality Trait: OPENNESS Amongst Groups

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.34700000	1.11566667	3.11	0.0384
Error	36	12.92700000	0.35908333		
Corrected Total	39	16.27400000			

R-Square	Coeff Var	Root MSE	op Mean
0.205665	13.08375	0.599236	4.580000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	3	3.34700000	1.11566667	3.11	0.0384

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	0.359083
Critical Value of Studentized Range	3.80880
Minimum Significant Difference	0.7217

Means with the same letter are not significantly different.				
Tukey Grouping		**Mean**	**N**	**group**
	A	4.8550	10	C
	A			
B	A	4.7150	10	E3
B	A			
B	A	4.6550	10	E2
B				
B		4.0950	10	E1

APPENDIX BB

ANCOVA Data Comparing Openness to Differences in Groups Self-assessment Scores

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	207.796966	51.949241	0.57	0.6835
Error	35	3168.922612	90.540646		
Corrected Total	39	3376.719577			

R-Square	Coeff Var	Root MSE	diff Mean
0.061538	256.7708	9.515285	3.705750

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	3	202.3635178	67.4545059	0.75	0.5325
op	1	0.1619382	0.1619382	0.00	0.9665

APPENDIX CC

Letter of Introduction

Researcher:
Trey Skaggs
UGA School of Music
250 River Rd.
Athens, GA 30609
C: 770-715-7282
tskaggs@uga.edu

Faculty Advisor:
Dr. Mary Leglar
T. 706-542-2755
mleglar@uga.edu

November 17, 2008

Dear Student,

I would like to ask your permission to videotape and allow other students to observe you while teaching lessons in class. Following the conclusion of each teaching episode, you will be given the opportunity to self-evaluate. The process of self-assessment can make identifying discrepancies in ones teaching much easier and help one become a better teacher.

If you agree to participate, you will be observed and videotaped as a part of your music education class. Videos of your teaching will be stored in a secure location and only released to you for viewing. Once viewed, the video will either be erased or returned to secure storage for the remainder of the study. After the study is completed, the video will be destroyed.

Your participation in this study is voluntary and not required as a part of this music class. It is my intent that this study will cause minimal disruption to your music education classroom. The video camera will be placed in the back of the room so it does not interfere with your lesson.

Self-assessment is an important aspect of teaching. Your participation in this study will help us to understand how we can better self-evaluate as music teachers and what changes if any are needed with current practices in teacher training. Thank you so much for your consideration.

If you have any questions about this study, you may contact Trey Skaggs or his supervising professor, Dr. Mary Leglar at any time. Again thank you for your consideration.

Sincerely,
Trey Skaggs

APPENDIX DD

Informed Consent

I agree to take part in a research study on self-assessment, which is being conducted by Mr. Trey Skaggs (770-715-7282) under the direction of Dr. Mary Leglar (706-542-2755) with the Music Education Department at the University of Georgia School of Music. I understand that participation is voluntary. I can refuse to participate or stop taking part at any time without giving any reason, and without penalty or loss of benefits to which I am otherwise entitled. I can ask to have information related to me returned to me, removed from the research records, or destroyed.

As a participant:

- I will complete a personality profile lasting thirty minutes.
- I will teach four twenty-minute lessons and fill out a self-assessment form after each lesson.
- My lesson will be videotaped and peers in the classroom will fill out an evaluation form at the conclusion of your teaching lesson.
- I will complete a five-minute peer critique form after other participants teach.
- All personal information will be kept confidential. Confidentiality procedures are as follows. A participation number will be randomly assigned with no personal data attached. For all forms and lessons you will be identified by this number rather than by your name or other personal data. All videotapes will be housed in a locked file cabinet that is within a locked office. You will only have access to recordings of yourself. At the conclusion of the study, you will have the option of keeping the recording of yourself only. If you choose not to keep their recordings, the recordings will be erased or destroyed.
- I will be required to spend no more than one hour per week in research activities.
- There is nothing that is expected to cause any harm or discomfort.
- Benefits of this study include: learning new ways in which to self-evaluate, being a part of a project that could find significant results, and learning how an experimental study functions at the graduate level. The potential benefit of this research for humankind is tremendous. All preservice and practicing teachers are evaluated each year. Hundreds of thousands of teachers receive poor teaching marks that they never expected. If effective methods can be devised that allow the preservice teacher to learn how to correctly evalute themselves, then the level of teaching should improve, there should be fewer poor teaching marks, and student achievement should improve. When student achievement improves, we have a greater number of quality contributing members to our society.
- I am entitled to contact the researcher at any time with questions or concerns. The researcher is available by telephone 770-715-7282 or by email tskaggs@uga.edu. You may also contact the supervising professor, Dr. Mary Leglar at 706-542-2755 or mleglar@uga.edu.
- I understand the study procedures detailed above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form to keep.

Trey Skaggs	_____	_____
Researcher	Signature	Date
_____	_____	_____
Name of Student	Signature	Date

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 612 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone 706-542-3199; email: IRB@uga.edu