DOCUMENT RESUME

ED 359 095 SO 022 804

AUTHOR Doane, Christopher; And Others
TITLE A Validation of Music Teacher Behaviors Based on Music Achievement in Elementary General Music Students.
PUB DATE 90
NOTE 11p.
PUB TYPE Journal Articles (080) -- Reports - Research/Technical (143)
JOURNAL CIT Research Perspectives in Music Education; v44 n1 p24-41 Fall 1990

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Educational Research; Elementary School Students; *Music Education; *Music Teachers; *Teacher Behavior; Teacher Evaluation; Teaching Methods
IDENTIFIERS *Florida

ABSTRACT

This purpose of the study was to examine music achievement of elementary general music students in relationship to teacher behaviors as measured by the Florida Performance Measurements System (FPMS), a research-based system for the evaluation of generic teacher behaviors and a Music Behavior Observation Form (MBOF), an investigator-devised form designed to assist in the assessment of typical musical behaviors required of teachers in elementary music classrooms. A total of 19 teachers and 1681 third grade students took part in the study, which took place in the elementary schools of a large west-central Florida county over the 1988-89 academic year. Student data collected included music achievement, pre- and post-instruction scores generated from a standardized county music assessment, music aptitude scores from the Gordon Primary Measure of Music Audiation, academic achievement assessments taken from CTBS (Comprehensive Tests of Basic Skills) scores in mathematics, reading and language and classification of students by socio-economic level, gender, gifted/SLD characteristics and ethnicity. Teachers were assessed over two 30-minute lessons by teams of trained observers using the FPMS and MBOF observation forms. Residualized achievement gain scores controlling for student characteristics were correlated with total scores and subscores on the FPMS and mean scores on the subtests and total MBOF. Analyses of the data revealed strong relationships between the teacher evaluation instruments. No significant relationships were found between teacher characteristics and residualized student achievement scores. Regression analyses showed that student achievement posttest scores could be predicted most effectively on the basis on pretest scores, rhythmic aptitude scores, and reading and mathematics achievement. Among the recommendations for further study were the need to expand the items that are contained on music achievement tests to more completely sample the diversity of learning that occurs in an elementary music classroom. (Author)

A VALIDATION OF MUSIC TEACHER BEHAVIORS BASED ON MUSIC ACHIEVEMENT IN ELEMENTARY GENERAL MUSIC STUDENTS

SD-AR

A VALIDATION OF MUSIC TEACHER BEHAVIORS BASED ON MUSIC ACHIEVEMENT IN ELEMENTARY GENERAL MUSIC STUDENTS

Christopher Doane, University of South Florida,
Claudia Davidsen, Hillsborough County Schools,
Jayne Hartman, St. Lucie County Schools

The purpose of the study was to examine music achievement in elementary general music students in relationship to teacher behaviors as measured by the *Florida Performance Measurement System (FPMS)*, a research-based system for the evaluation of generic teacher behaviors and a *Music Behavior Observation Form (MBOF)*, an investigator-devised form designed to assist in the assessment of typical musical behaviors required of teachers in elementary music classrooms. A total of nineteen teachers and 1681 third grade students took part in the study which took place in the elementary schools of a large west-central Florida county over the 1988-89 academic year. Student data collected included music achievement pre and post instruction scores generated from a standardized county music assessment, music aptitude scores from the Gordon Primary Measures of Music Audiation, academic achievement assessments taken from CTBS scores in mathematics, reading and language and classification of students by socio-economic level, gender, gifted/SLD characteristics and ethnicity. Teachers were assessed over two 30 minute lessons by teams of trained observers using the FPMS and MBOF observation forms.

Residualized achievement gain scores controlling for student characteristics were correlated with total scores and subscores on the FPMS and mean scores on the subtests and total MBOF. Analyses of the data revealed strong relationships between the teacher evaluation instruments. No significant relationships were found between teacher characteristics and residualized student achievement scores. Regression analyses showed that student achievement posttest scores could be predicted most effectively on the basis of pretest scores, rhythmic aptitude scores, and reading nd mathematics achievement. Among the recommendations for further study were the need to expand the items that are contained on music achievement tests to more completely sample the diversity of learning that occurs in an elementary music classroom.

INTRODUCTION AND STATEMENT OF THE PROBLEM

The validation of effective teaching behaviors is a major concern of everyone involved in education. Teacher training institutions have the responsibility to train prospective teachers and to seek confirmation that the training these future teachers receive will be effective in later helping to produce students who achieve at a high level. Schools at all levels carry a clear mandate to insure that the instruction being delivered is efficient and effective in meeting the needs of the student population. These powerful ideas lead to the consideration of very important issues facing the profession in such areas as curriculum development, teacher evaluation, and educational accountability.

Among the many responsibilities of the education community is the evaluation of the quality of learning that takes place in the multitude of instructional settings that exist in the modern educational environment. Systematic efforts in this evaluation focus primarily on two areas; first, the particular teaching behaviors of the teacher and second, the specific, measurable learning that takes place by the student. The proliferation of teacher evaluation instruments occurring over the past several years, as well as the public's preoccupation with standardized achievement test scores, point to the importance placed on evaluation in today's environment in the schools. Notwithstanding the well-publicized discussion of these issues, the fact remains that the consideration of the influence of specific behaviors of teachers on student learning outcomes constitutes one of the most important questions facing the teaching profession today. It is only through establishing connections between teacher behaviors and their potential effects on student learning that teaching behaviors become valid.

In such an evaluation, issues relating to the reliability and validity of the measurement situation are of paramount importance. While some aspects of the classroom environment can be controlled to a high degree, there are still significant aspects of the evaluation process that remain unpredictable. Among these factors are the changing and problematic nature of the day to day classroom environment, the consistency of teaching behaviors, and issues relating to the quality of observation instruments and the reliability of the teacher evaluator. These and other factors should all introduce a degree of caution to the observer when interpreting the results of teaching evaluations.

Across the education profession, approaches to establish the validity of teaching behaviors have taken many different directions. Content validity of behaviors has been established largely through a consensus of professionals. Other approaches have attempted to establish criterion-based validity of behaviors by seeking a relationship between specific teaching behaviors and student learning based on a specified measure of achievement. To date, efforts in this particular direction have been made in many geographical locations throughout the country and in a wide variety of subject areas.

In counties throughout Florida, teacher behaviors are monitored through the *Florida Performance Measurement System (FPMS)*, a measurement system of generic teacher behaviors developed in Florida in the early 1980's. The FPMS instrument is currently used in cross disciplinary subject areas as a means of evaluating first-year teachers in the Beginning Teacher Program in Florida and is employed in other teacher evaluations on a systematic and regular basis. Although FPMS is recognized throughout the state as a valid and reliable means of assessing teaching behaviors to date, the instrument has not been validated in an arts-related content area.

Purpose of the Study

The purpose of this study is to examine the relationship of teaching behaviors utilized by elementary general music teachers and student achievement in general music. Considerations in this examination deal with the role of specific music behaviors that the teacher demonstrates and how these music behaviors impact learning in students with differing levels of music ability and from a wide range of social and economic backgrounds. Specific questions to be investigated include:

1) Does music teacher performance as measured by the Florida Performance Measurement System correspond to student achievement in elementary music classes?

2) Does the accuracy of music behaviors in the classroom as demonstrated by the teacher have a significant effect on the amount of musical learning that occurs in the classroom?

3) What is the interaction between music aptitude, music achievement and music teacher characteristics?

Consideration of these questions may add further validity to the use of the Florida Performance Measurement System as a cross-disciplinary predictor of student achievement and will increase the knowledge base currently in place with regard to effectiveness in music teaching.

Related Literature

There have been essentially two ways that investigators have approached the problem of defining the characteristics of an effective teacher. One way has been to consult experts in education, asking them simply to describe the characteristics that they themselves feel are the most crucial for success. A second way has been to relate teacher attributes to student achievement or affective response. Although the evidence is not conclusive, some reports indicate that the two means of examining the question are not mutually supportive (Coker, Medley and Soar, 1980; Silvernail, 1979). Furthermore, Dunkin and Biddle (1974) cited a number of studies dealing with classroom climate, teacher directiveness, discipline and group management, and behavior modification in which conflicting or curvilinear/complex findings were reported. Coker, Medley and Soar (1980) reported non-significant and/or negative correlations between teacher characteristics generally regarded as effective and measures of student achievement or attitude.

Berliner (1976) examined several problems associated with the issue of relating teacher "process" to educational "product". Among these were the validity of the achievement measure, the appropriateness of measuring short term effects, the limited ability of the test measures to examine a wide universe of the students' experience the stability of the teachers' teaching behaviors, and the evaluation problem of determining the validity of the demonstrated teaching behaviors to the educational task in the classroom.

Several authors have reported on teacher attributes associated with enhanced student achievement (Bennett, 1976; Brophy and Good, 1974; Medley, 1977; Medley and Crook, 1980; Olivia and Henson, 1980; Rosenshine and Furst, 1971). Studies in music that identified desirable music teacher attributes were reported by Doane (1983), Edwards (1979), Froelich (1977), and Taebel (1980). Efforts to validate these behaviors in light of student achievement were examined in studies by Edwards (1979), Froelich (1977, 1979) and Reid (1972). Of the process-product research completed in music education, it is appropriate to focus on the findings of a study by Taebel and Coker (1980) because of this study's close relationship to the present study. A total of 26 competencies were identified for study to determine if a correlation existed between these attributes, attitude gains, and/or student achievement (as measured by a nationally standardized achievement test in music). Three competencies correlated at moderate levels with achievement. Only one competency, "teacher relates lesson objectives to student interests and needs", related significantly to achievement. The only characteristic to relate significantly to both achievement and attitude gains was "students initiate verbal interaction".

The inconsistent results posed by a number of existing studies point to the fact that an inventory of musical and professional teaching behaviors that lead to more effective music teaching as defined by enhanced student achievement has not yet been identified.

Methodology

The settings for the study were selected third grade classrooms in the elementary schools of a large urban school district of west central Florida. A total of nineteen elementary schools were selected for the study to create a balanced sample based on geographical, socio-economic, and academic achievement levels representative of the eighty elementary schools in the county. Additionally, nineteen teachers were selected for the study based on their teaching assignment at the targeted schools as well as their teaching and performance backgrounds that would provide a representation of the entire elementary music faculty of 134 teachers. Criteria used to select the teachers included gender, number of years of teaching experience, and major performance instrument. No beginning teachers were selected for the study and, in schools where there were multiple teachers responsible for instruction, a teacher was selected to take part in the study only if the students in a classroom to be evaluated were a part of the regular teaching assignment of the teacher. Of the nineteen teachers who took part in the study five were male and fourteen were female. Seven teachers considered voice to be their primary instrument while nine were performers on another type of instrument. Three teachers reported equal abilities on both voice or another instrument. In terms of teaching experience, four teachers had from two to five years experience, three teachers from six to ten years, eight teachers from eleven to nineteen years experience, and four teachers more than twenty years of experience.

The 1681 students taking part in the study were members of intact third grade classrooms in each of the target schools. Complete data sets were completed for 1263 students. Several data collection points occurred throughout the year. Student absence from school accounted for a portion of the incomplete data while students transferring into the school during the 1988-89 academic year accounted for incomplete academic achievement (CTBS) records for some of the students.

Description of the measurement instruments

Florida Performance Measurement System

One of the principal means of gathering information on specific teaching behaviors was through the FPMS Summative Evaluation Form. While complete details concerning the development of this instrument are readily available, a brief description of the instrument may be helpful. The Florida Performance Measurement System was developed in the early 1980's after a mandate from the Florida state legislature and the Department of Education. After an extensive survey of the process-product research in teacher effectiveness, a series of generic (non-subject specific) teaching behaviors were selected that were supported by a significant body of research as being associated with enhanced student learning. Four specific areas or domains were identified that could be assessed by a trained classroom observer and incorporated into a Summative Evaluation Form. This observation instrument was used in the present study to assess the domains of instructional organization and development, presentation of subject matter, communication and class management.

The FPMS has been extensively submitted to tests of content validity and reliability (Teacher Evaluation Project, 1982-1983), normed on a representative sample of Florida teachers (Teacher Evaluation Study, 1983-1984), and tested in several studies for predictive validity (Teacher Evaluation Studies, 1984-1985 and 1985-1986).

Music Behavior Observation Form

The Music Behavior Observation Form was specifically designed for this study to measure musical behaviors related to music teaching activities in an elementary general music classroom. Comprehensive categories of musical behaviors were developed by a panel of experts consisting of university music education professors and public school music supervisors. These behaviors were selected based on an assessment of those activities most valued in general music lessons in the county. Behaviors selected were music performance skills, planning and leading movement activities, planning and presenting listening activities, and demonstrating desirable classroom interaction behaviors. Performance skills were further defined as singing proficiency, rhythmic skills, and performing on accompanying instruments. Interaction behaviors were grouped into organization skills, explanation skills, and response to musical errors.

A series of six teaching episodes were videotaped utilizing elementary music teachers not participating in the study. Of these lessons, three were scripted and three were not scripted. Teachers to be filmed were instructed to include accurate as well as inaccurate musical behaviors in their lessons. These teaching episodes were later evaluated by a panel to determine the stability and effectiveness of the instrument and to refine the definitions of the behaviors to be observed. After further refinement of the instrument, training sessions utilizing four of the teaching episodes were conducted with five experts who were to act as music behavior evaluators. In this group were three music education professors, an adjunct music faculty from a local university, and a music supervisor from a nearby county. Interjudge reliability between these observers was calculated to be .91 using a total agreement/agreement + disagreement formula. Also present at the training session were observers who were to evaluate teachers using the FPMS Summative Observation Form. These people were present to review the types of activities in which music teachers engage and to clarify aspects of the lesson that might have been unfamiliar.

In using the Music Behavior Observation Form the observer tallies accurate (appropriate, effective) teacher behaviors as well as inaccurate (inappropriate, ineffective) behaviors during the lesson and uses these measurements as a means of assigning an overall rating of 1 (ineffective) to 4 (effective) to the major areas of Singing, Rhythm Skills, Accompanying, Overall Performance, Movement, Listening, Organization, Explanation, and Response to Errors categories. Observers tally the non-presence of any of these behaviors by coding a zero in the appropriate category.

Primary Measures of Music Audiation

The Primary Measures of Music Audiation (Gordon, 1979) is a standardized music aptitude test designed for use by students in kindergarten through the third grade to assist in measuring students' ability to internally conceptualize and compare basic tonal and rhythm patterns. The PMMA is a tape recorded group test divided into Tonal pattern and Rhythm pattern subtests, each subtest consisting of 40 items. Each subtest must be administered separately with a total administration time of 20 minutes for each testing period.

The initial research into the development of the PMMA was begun in 1971. Over the next seven years a taxonomy of tonal and rhythm patterns was identified and validated by grouping each pattern into items of relative difficulty. The Primary Measures of Music Audiation includes those patterns that were determined to be in the category of easy patterns to hear and retain. Subsequent research with third grade students revealed split-half reliability estimates of .70 for the Tonal test, .64 for the Rhythm test, and a composite reliability coefficient of .72. Correlations with estimates of reading and mathematics, readiness to enter school, or gender suggested no systematic relationship with these factors and implied that the PMMA was independently measuring a type of music aptitude County Music Assessment (Level II).

The *measure of musical achievement* was provided by the County Music Assessment, level II (grade 3). The assessment is used throughout county elementary schools as a pre- and post-instruction means of assessing the progress of students and as a means to refine curricular offerings.

The 41 test items draw on knowledge that is primarily cognitive in nature. The basis for item selections were the goals, objectives and assessment guidelines identified in the Florida Catalogue of Music Objectives (1974) and the County Elementary Music Expectancies (1984). Content validity of the assessment was established by deriving the test items from materials drawn from the County Curriculum Guide and the Florida Catalogue of Objectives. Reliability of the assessment is estimated to be .75 based on the KR-20 means of calculation. Components of the CMA are: 1) a cassette tape on which musical examples and questions for both

aural and visual sections are recorded, 2) test booklets of questions and multiple choice options featuring verbal and pictorial responses, and 3) an answer sheet that may be scored by machine. The total time necessary to administer the test is thirty minutes.

Procedures

During September of the 1988-89 school year participating music teachers administered the Primary Measures of Music Audiation to their intact classes of third grade students. The following week the County Music Assessment was given to all students in the participating classes. Following the testing period music teachers forwarded all test results to the County Office of Testing and Evaluation for recording. Teachers were not informed of student performance on the music aptitude measure, but were advised concerning student performance on the assessment pretest in order to plan their instructional activities most effectively. During the fall additional data was collected concerning student characteristics of gender, ethnic background, participation in gifted or learning disability programs, socio-economic status as determined by participation in various levels of reduced cost lunch programs, and academic achievement based on CTBS scores in language, mathematics, and reading.

Participating teachers followed their own teaching plans throughout the balance of the school year. Instructional time was equivalent in all the participating classrooms. In March and again in April-May participating teachers were evaluated by an observation team consisting of a state-certified FPMS evaluator and one of five trained music observers. Each member of the team viewed a thirty minute lesson and independently scored the lesson without consultation. The teachers under observation were informed of the visit prior to the class period and were asked to prepare a lesson plan, employ their customary teaching behaviors, and to include a variety of activities in the lesson. After the observation period, teachers were not consulted concerning their performance in the lesson and the observers were instructed to maintain confidentiality with regard to their reactions to the teacher or the lesson. In May of 1989 all third grade classrooms in the county were administered the Music Assessment test. Tests were scored in individual schools and results were forwarded to the Office of Testing and Evaluation where the results were recorded along with other data for examination.

Presentation of the Data

A summary of the evaluations of the teacher behaviors based on scores derived from the *Florida Performance Measurement System (FPMS)* Summative Form and the *Music Behavior Observation Form (MBOF)* is given in Table 1. The range of possible scores on the FPMS instrument is from 40-100 for the total instrument which can be grouped into a range of 20-60 for effective behaviors and 20-40 for ineffective behaviors. Mean scores are given for the MBOF. A score of 1 indicates a behavior judged to be inadequate while a score of 4 indicates an effective behavior. Mean scores for components of the MBOF range from 2.92 to 3.55, indicating behaviors which may be interpreted as typically adequate to effective behaviors for teachers taking part in the study. Variability indicated by the range of scores suggests that the observers and the instrument were sensitive to differences among teachers and behaviors displayed in the classrooms.

Table 1

Descriptive Statistics for Teacher Performance Measures

	Range	Mean	S.D.
FPMS			
Effective	39.5 - 53.0	44.29	2.16
Ineffective	30.0 - 38.0	34.95	1.89
TOTAL	69.5 - 89.5	79.24	4.42
MBOF			
Singing	1 - 4	2.97	.84
Rhythm	1 - 4	3.47	.70
Accompanying	2 - 4	3.41	.83
Perf. Overall	2 - 4	3.28	.71
Movement	1 - 4	2.93	.89
Listening	3 - 4	3.55	.81
Organization	1 - 4	3.16	.76
Explanation	2 - 4	3.18	.75
Responses to Error	1 - 4	2.92	.82
AVERAGE	1.79 - 4.0	2.95	.73

Correlations between components of the FPMS and MBOF evaluations are presented in Table 2. Of particular interest in this analysis are the significant and positive intercorrelations between several components on the MBOF instrument as well as the significant, positive correlations between the FPMS effective behaviors category and six of nine components of the MBOF. Additionally, there were positive, but not statistically significant correlations between the ineffective behavior component of the FPMS and MBOF components. Generally, this analysis demonstrates the internal consistency of evaluation of particular categories of behaviors on the MBOF and the consistency between evaluations utilizing FPMS and MBOF.

Table 2

Correlations Between FPMS and the Music Behavior Observation Form (MBCF)

	FPMS			Music Behavior Observation Form (MBOF)									
				Performance				Movement	Listening	Classroom Interaction			
	Effective	Ineffictive	Total	Singing	Rhythm	Accom-panying	Overall			Organization	Exploration	Response to Error	Average
FPMS													
Effective	1.00												
Ineffective	.64*	1.00											
Total	.94**	.85**	1.00										
MBOF													
Singing	.47*	.24	.41	1.00									
Rhythm	.43	.29	.41	.76**	1.00								
Accompanying	.59*	.37	.55*	.76*	.71**	1.00							
Perf. Overall	.52*	.33	.49*	.94**	.89II	.81**	1.00						
Movement	.41	.39	.44	.75**	.81**	.86**	.86**	1.00					
Listening	.43	.29	.39	.76**	.71**	.81**	.84**	.85**	1.00				
Organization	.54*	.39	.53*	.46*	.66*	.73**	.63**	.82**	.74**	1.00			
Exploration	.48*	.36	.48*	.61*	.75**	.68**	.79**	.82**	.75**	.67*	1.00		
Response to Error	.54*	.45	.56*	.48*	.66*	.79**	.64*		.77**	.68*	.76**	.69*	1.00
AVERAGE	.54*	.35	.51*	.83**	.82**	.89**	.91**	.95**	.96**	.81**	.84**	.71*	1.00

* $p < .05$; ** $p < .001$

A summary of student performance measures included an examination of student music aptitude as determined by the PMMA subtests in tonal and rhythm aptitude and music achievement levels based on scores on the pre and posttest administrations of the *County Music Assessment*. Table 3 illustrates the results of these measures. Of particular interest is the increase in the mean score and diversity of scores on the post-instruction achievement measure. Further examination of this area using a regression analysis indicated that pretest scores were the best predictors of posttest scores and that rhythmic aptitude, reading and mathematics achievement were also significantly predictive ($p < .05$) of posttest scores when other student variables were controlled. Tonal aptitude and language achievement were not significantly related.

Table 3

Descriptive Statistics for Student Measures

	School Range	Student Mean	S.D.
ACHIEVEMENT (0-41)			
Pretest	18.75 - 23.13	20.57	1.26
Posttest	18.61 - 29.58	24 35	2.50
APTITUDE (0-41)			
Rhythm	31.15 - 33.74	32.43	3.56
Tonal	32.13 - 35.69	33.54	4.02

The relationship of teacher behaviors, music aptitude, and music achievement as measured by the FPMS, MBOF, PMMA, and CMA respectively was determined by a series of multiple correlations utilizing mean posttest scores. These relationships are illustrated in Table 4.

Table 4

Correlations between Student Music Achievement, Student Aptitude

	Student Achievement		Aptitude		Teacher Performance FPMS		
MBOF Average	Pre	Post	Rhythm	Tonal	Eff	Ineff	Total
ACHIEVEMENT							
Pretest	1.00**						
Posttest	.42**	1.00**					
APTITUDE							
Rhythm	.27**	.27**	1.00				
Tonal	.16**	.08**	.26*	1.00**			
FPMS							
Effective	.03	-.14		1.00			
Ineffective	.07	-.07		.64**	1.00		
TOTAL	.05	-.13			.94**	.85**	1.00
MBOF							
Average	-.38	-.41		.54*	.35	.51*	
1.00							

* $p < .05$
** $p < .01$

In order to account for possible differences in achievement scores attributable to student characteristics, residualized gain scores were used as a means for controlling for differences in sample characteristics. A residualized gain score is the difference between the obtained posttest score and the expected posttest score predicted on the basis of a particular variable or linear combination of variables. A regression model was used to generate expected posttest scores while controlling for student characteristics including: academic achievement (as defined by CTBS scores in mathematics, language, and reading), socioeconomic status (as defined by levels of participation in a subsidized lunch program), ethnic characteristics, pretest scores, participation in gifted or learning handicapped programs, and music aptitude as measured by PMMA scores on tonal and rhythmic subtests. A series of student residualized gain scores on the *County Music Assessment* was correlated with teacher performance as measured on subtests of the FPMS and MBOF to determine whether particular teacher behaviors might have a differential effect on students with diverse characteristics. These relationships are shown in Table 5. Given the relationship of FPMS evaluations to those of the MBOF (see Table 2) and the non-significant relationship of FPMS scores to student achievement, the negative, but predominantly non-significant correlations were expected. Further examination of teacher behaviors compared to student scores on particular musical concepts contained in the CMA including cognitive, aural, and notational information revealed similar patterns of non-significant relationships.

Table 5

Correlations between Teacher Behaviors and Residualized Student Achievement Scores

Teacher Behaviors	Academic Achievement	Pre-Test Score	Ethnicity Gifted SLD/SES	Music Aptitude
FPMS				
Effective	-.21	-.18	-.23	-.32
Ineffective	-.13	-.11	-.15	-.24
Total	-.19	-.16	-.22	-.31
MBOF				
Total	-.15	-.33	-.24	-.06
Singing	-.07	-.16	-.07	-.02
Rhythm	-.40	-.45	-.20	-.13
Accompanying	-.21	-.28	-.19	-.08
Ocverall Explanation	-.34	-.51*	-.39	-.26
Response to Errors	-.36	-.46	-.24	-.15

* $p < .05$

Schools were ranked in descending order by differences between mean pretest and posttest scores on the music achievement test in order to determine if the teachers in highly ranked schools displayed behaviors significantly different from those teachers in the lower ranked schools. Table 6 shows the student and teacher data for the five highest ranked and lowest ranked schools. It is interesting to note that the highest ranking schools based on a mean difference score also had achievement posttest scores that were significantly higher than those of the lowest ranked schools, suggesting a true difference in achievement between these schools. However, based on FPMS scores, the teachers in the lowest ranking schools were slightly more effective than the teachers in the highest ranked schools. Other points to note are that the highest teacher score on FPMS and the highest score on the MBOF were earned by the teacher at the fifth-ranked school. The lowest score on the FPMS was assigned to the teacher at the top-ranked school and the lowest score on the MBOF was earned by the teacher at the second-ranked school. There were no apparent differences between student populations in the two groups of schools based not only on music aptitude, as indicated in Table 6, but also in proportions of gifted/SLD, high/low socioeconomic level students, or ethnic representation.

Table 6

Comparisons of Teacher Performance Measures for Highest and Lowest Ranked Schools

RANK	Music Achievement Pre	Music Achievement Post	Aptitude Rhythm	Aptitude Tonal	FPMS Effective	FPMS Indirective	FPMS Total	Music Behavior Observation Form (MBOF) Performance Singing	Performance Rhythm	Performance Accompanying	Overall Movement	Listen Organization	Explanation	Classroom Interaction Response to Error	Classroom Interaction Average
1	21.6	29.6	32.5	34.2	41.5	32.5	74.0	2.5	3.0	3.0	2.5	3.0	3.5	2.0	2.5	2.72
2	21.9	28.1	33.7	33.8	43.5	34.0	77.5	3.0	3.0	2.0	.25	1.0	3.0	3.0	2.6	2.33
3	19.6	25.5	32.4	32.7	41.0	34.5	75.5	3.5	3.5	3.5	3.5	0.0	4.0	2.5	2.5	3.19
4	19.2	24.5	32.2	32.1	44.5	34.5	79.0	3.0	4.0	3.0	3.5	3.0	4.0	4.0	4.0	3.56
5	19.4	24.3	32.6	33.8	46.0	35.0	81.0	4.0	4.0	4.0	4.0	0.0	4.0	4.0	4.0	4.00
15	20.1	22.7	32.6	32.4	44.5	35.0	79.5	4.0	4.0	4.0	4.0	4.0	3.0	3.5	3.0	3.69
16	20.8	23.1	32.5	32.0	45.0	34.5	79.5	3.0	3.5	3.0	3.5	3.0	3.5	4.0	2.5	3.22
17	18.7	20.4	31.3	32.7	44.0	39.5	79.5	2.0	4.0	4.0	3.0	4.0	4.0	4.0	4.0	3.62
18	21.5	22.8	33.3	33.8	44.0	35.0	79.0	3.0	2.5	3.0	3.0	2.5	3.0	2.5	2.5	2.75
19	19.4	18.6	32.0	33.2	45.0	34.0	79.0	3.5	4.0	4.0	4.0	4.0	4.0	4.0	4.0	3.94

Note: Score of 0 on MBOF components means that there was no opportunity to observe the behavior.

insert Table 6 here

In order to determine whether individual indicators in the FPMS system might be related to levels of student achievement, a series of t-tests were computed between teacher scores form the five highest and lowest ranking schools. Additionally, item agreements for each of the forty indicators were computed in order to determine the relative stability of the teacher behaviors over the two FPMS observations. Individual item agreements ranged from .43 to 1.00 with a mean of .75. Results of the t-tests indicated a statistically significant ($p < .05$) difference in scores when compared to achievement at the ranked schools only for the indicator "emphasizes important points". No other individual indicators from FPMS were related to student achievement.

When further examination of the teacher data revealed no apparent systematic differences between teacher behaviors an examination of teacher background variables was undertaken. These factors, including gender, years of teaching experience and music performance major (voice or instrument) were compared to student achievement using an ANOVA procedure. No significant relationships were found between performing instrument and student achievement. However, there was a difference in student achievement based on the number of years of teaching experience of the classroom music teacher. Significant ($p < .05$) F values were found between all residualized student achievement measures and teachers who had more than 20 years of teaching experience. No other statistically significant relationships were detected.

Summary,Discussion, and Conclusions

The first major question under examination in the study dealt with the relationship of music teacher performance as measured by the *Florida Performance Measurement System Summative Evaluation Form* and student achievement in third grade music classrooms as measured by the *County Music Assessment (Level II)*. An examination of this relationship showed that there was no significant relationship between student achievement and FPMS teacher evaluations when student posttest scores were predicted

from pretest scores, music aptitude, demographic variables such as gender, race, socio-economic level or gifted/SLD status, or achievement attributes based on CTBS scores in mathematics, language and reading. The individual teacher attribute "emphasizes important points" was significantly related to student achievement at both the five highest and five lowest ranking schools. No other individual attribute was related to musical achievement.

The second research question related to the accuracy of teacher music behaviors and the possible effect of this accuracy on student achievement. As in the case of FPMS evaluations, there were no significant relationships between any musical or organizational skills as measured by the *Music Behavior Observation Form* and student achievement on the *County Music Assessment (Level II)*. In an additional observation, the evidence from this study suggested that the number of years of experience in the teacher's background may be related to student achievement. Due to the relatively small number of teachers in various categories of experience levels who took part in this study this factor should be examined in additional research efforts.

The third question examined the interaction between music aptitude, music achievement and music teacher characteristics. Rhythmic aptitude as determined by student performance on the *Primary Measures of Music Audiation* was positively and significantly correlated with achievement on both the pretest and posttest at the .27 level. Tonal aptitude was positively related to achievement on the pretest at the .16 level and the posttest at the .08 level. Teacher behaviors as measured by the *Florida Performance Measurement System Summative Evaluation Form* or the *Music Behavior Observation Form* were not related to student achievement residualized gain scores controlling for music aptitude.

The results of the study illustrate and confirm the difficulty in determining the relationships between teacher behaviors and changes in student achievement. While there is evidence from this and other studies to suggest that student attributes such as academic achievement, music aptitude and socio-economic level do play a significant role in predicting levels of achievement in music, there appeared to be no systematic relationship between these student attributes and categories of teacher behaviors that were examined. Although not statistically significant, the low and/or inverse relationships between commonly supported teacher behaviors and student achievement is both provocative and disturbing.

However, just as the profession should attempt to validate particular teaching behaviors, it as also necessary to validate the measurement instruments by which definitions of learning are made. In the present study, there is no question that the achievement test instrument, the County Music Assessment, is both reliable and valid for the instructional program in the target schools. As such, it does represent a useful tool for which it was designed, that is, a means by which teachers may assess the progress of students and to plan for curricular change. However, as the education profession is learning in many other instructional situations, single measures of learning in complex areas of study are seldom adequate to measure the variety of types of learning that may take place in the classroom.

Most music teachers are concerned not only with cognitive categories of learning, but also with individual musical skill development and the continued development of positive attitudes toward music participation. Many educators might argue that cognitive aspects of learning in music may be the least important priority for elementary students. Given this argument, there is ample justification to suggest that there are many broad categories of learning that occur in successful music classrooms and that a single measure, such as that used in this study, is neither sensitive enough nor sufficiently comprehensive to provide a complete assessment of musical achievement or the success of the teacher. Future studies of this type should seek to incorporate measurement instruments that are sensitive to a broad and comprehensive inventory of musical and non-musical learning that may occur. Among these areas to consider are individual and group performance growth in music, the development of self-esteem and attitude in music class, and the selection of appropriate cognitive areas of musical growth.

Generic measures of behaviors such as the *Florida Performance Measurement System* certainly have their place in the systematic evaluation of teachers. Other studies in music have illustrated that within such systems it is possible to extract particular definitions of behavior that indicate significant relationships to student learning. Given the usually small to moderate relationships that exist in these instances it seems reasonable to focus on validating individual behaviors within the larger evaluation system with a particular view toward building a comprehensive evaluation system incorporating a variety of generic as well as specific musical behaviors.