2000

Factors Related to the Validity of Reproduction Tonal Memory Tests

Charles E. Norris
Grand Valley State University, norrisc@gvsu.edu

Follow this and additional works at: https://scholarworks.gvsu.edu/mus_articles

Part of the [Music Commons](#)

ScholarWorks Citation

Norris, Charles E., "Factors Related to the Validity of Reproduction Tonal Memory Tests" (2000). *Peer Reviewed Articles*. 2.
https://scholarworks.gvsu.edu/mus_articles/2

This Article is brought to you for free and open access by the Music & Dance Department at ScholarWorks@GVSU. It has been accepted for inclusion in Peer Reviewed Articles by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

The purpose of this investigation was to acquire data regarding the validity of reproduction tonal memory tests by examining the relationships between performances on an existing reproduction tonal memory test and performances on several recognition tonal memory tests. Analysis of 210 fifth- through twelfth-grade students' scores on one reproduction test and four recognition tests of tonal memory revealed moderate relationships ($r = .47$ to $.66$) between the two types of tests. Stepwise multiple regression revealed a moderately strong predictive relationship between a combination of three recognition measures and the reproduction test ($R = .73$, $R^2 = .53$). Little change was noted in the relationship between the two types of tests with regard to singing experience. Although the reproduction tonal memory test was tentatively deemed a valid measure, the following additional research and analysis was suggested: (a) continued consideration of gender, ability, and total musical experience as they affect overall validity, and (b) the creation of a more reliable recognition test alternative.

Charles E. Norris, *Grand Valley State University*

Factors Related to the Validity of Reproduction Tonal Memory Tests

Commonly prescribed measures in predicting success in choral music ensembles are reproduction tonal memory tests, in which multi-pitch patterns are played on an instrument and the subject sings back what was heard. Authors of choral methods textbooks not only advocate such a practice, but also maintain that reproduction tonal memory tests can help music directors evaluate students' overall aural skills, ability to learn new choral material, and the potential for sight-singing achievement (Brinson, 1996; Lamb, 1986; Roe, 1983).

Although it may be assumed that reproduction tonal memory tests are commonly used, there has been little psychometric data that supports the validity of this type of test, suggesting the likelihood that

Charles E. Norris is coordinator of music education at Grand Valley State University, 1 Campus Drive, Allendale, MI 49401-9403; e-mail: norrisch@river.it.gusu.edu. Copyright © 2000 by MENC—The National Association for Music Education.

such tests are often administered and scored under conditions of little control and rigor. Virtually no research on the validity of reproduction tonal memory tests exists, with the exception of a single study (Mowrer, 1996) involving the prediction of an individual's ability to contribute to a choral ensemble based on the individual's score on a reproduction tonal memory test. By correlating scores of performances on his seven-item tonal memory test with teacher ratings of performances in choral settings, Mowrer was able to claim that his test was valid ($r^2 = .21-.47$ in five different choirs). Reliability was also established using the test-retest method ($r = .86$).

Further concern over the validity of reproduction tonal memory tests may be noted when one considers that what one does in reproduction tonal memory tests could be substantially different than what one does in recognition tests of tonal memory (Bartlett, 1996). The recognition tests or paper-and-pencil tonal memory tests generally require the subject to make discriminations between two sets of tonal patterns. For instance, the tonal imagery section of Gordon's Musical Aptitude Profile (1965) requires listeners to determine whether the pitch materials of two stimuli are the same or different. The Seashore tonal memory subtest (Seashore, Lewis, & Saetveit, 1960) further challenges one by requiring a determination of which pitch in two consecutive patterns has been altered. Bartlett argues that reproduction tonal memory tests are more difficult because they involve not only aural discrimination skill (recognition), but also the ability to replicate what is perceived via the singing voice (reproduction). Considering Bartlett's proposition and Rutkowski's (1990) observation that many existing measurement devices designed to assess vocal ability also measure tonal skills, it is possible to assume that tests intended to measure tonal skills via the singing voice may also be measuring vocal development and/or singing ability.

Various studies related to singing ability trace the development of vocal pitch-matching accuracy (e.g., Davies & Roberts, 1975; Geringer, 1983; Goetze, Cooper, & Brown, 1990), indicating that as children mature, vocal pitch-matching accuracy increases. In other studies, researchers have demonstrated that model of presentation (sine-wave, child voice, female voice, male voice, falsetto, and presence of vibrato) and octave may have an effect on the pitch-matching accuracy of both inaccurate and accurate singers (Price, Yarbrough, Jones, & Moore, 1994; Yarbrough, Bowers, & Benson, 1992; Yarbrough, Green, Benson, & Bowers, 1991; Yarbrough, Morrison, Karlick, & Dunn, 1995).

The relationship between pitch-matching accuracy and aural discrimination has also been investigated. Bentley (1969) determined that the performance of inaccurate singers was weaker on pitch-discrimination tasks (paper-and-pencil tests) than was the performance

of their accurate-singing counterparts. Supporting Bentley's findings were studies conducted by Pedersen and Pedersen (1971) and Zwissler (1971), who found moderate relationships between the abilities to discriminate between two sets of pitches in the recognition sense and to represent what was heard through the singing voice. Geringer (1983) found a moderate relationship between vocal pitch-matching accuracy and aural discrimination skills in high-ability fourth graders. Contrary to the aforementioned findings, investigations by Cooper (1992) and Rutkowski (1996) have yielded data that indicate a weak, if any, relationship between vocal pitch-matching accuracy and aural discrimination skills.

The lack of psychometric data on the validity of the reproduction tonal memory test, questionable representation of tonal memory via the singing voice, and ambiguous data regarding the relationship between use of the singing voice and aural discrimination skills have provided the impetus for the principal questions of this study. This inquiry is also meant to further research efforts on vocal pitch-matching accuracy, aural discrimination, tonal memory, or any combination thereof beyond upper elementary years.

Purpose

The purpose of this investigation was to acquire data about the validity of reproduction tonal memory tests by examining the relationships between performances on an existing reproduction tonal memory test and performances on several recognition tonal memory tests. Examination of this relationship may assist the choral music education community in making decisions on the validity of reproduction tonal memory tests and their usefulness in evaluating student potential. Of additional concern is the extent to which the relationship between the two types of tonal memory tests varies with singing experience. A final intention is to determine if there is one recognition test or a combination of recognition tests that can adequately predict one's reproduction tonal memory. Such a test or tests might be an even more reliable and valid alternative to what is considered by the choral music education community to be a highly useful diagnostic resource.

METHOD

Design and Selection of Subjects

A cross-sectional design was used so that the relationship between several recognition tests and one reproduction test of tonal memory could be observed over a span of developmental stages. The sample

was comprised of 5th- through 12th-grade students from one school in the midwestern United States that has maintained the same music teaching faculty for 13 years. Students who participated in the program had the benefit of consistent instructional content and style throughout their musical careers. The instructional program in music at all levels has been noted for excellence as evidenced by consistent high achievement of its students at interscholastic music competitions at all grade levels. Furthermore, the program is highly regarded by music education specialists in its immediate area. In consideration of Gordon's (1984) assertion that musical aptitude becomes stabilized at around age 9, and acknowledging that most organized school choral activities begin in Grade 5, I set out to study 30 choral music students in each of Grades 5, 6, 7, 8, 9, 10, 11, and 12. A request for subjects who had continuous enrollment in the general music and vocal music programs since entering kindergarten yielded 210 volunteers divided as follows: Grade 5—29, Grade 6—23, Grade 7—29, Grade 8—27, Grade 9—27, Grade 10—27, Grade 11—25, and Grade 12—23.

Data Collection

The data collected were raw scores on four recognition tests of tonal memory and one reproduction test of tonal memory. Selection of three recognition tests was based on studies indicating that presence or lack of tonality, length, and presence of rhythm affect perception and retention of musical stimuli (Boisen, 1981; Long, 1977; Sink, 1983; Pembroke, 1986; Taylor, 1976). Selected tests were Seashore Measures of Musical Talents (Seashore, Lewis, & Saetveit, 1960), tonal memory subsection (Series B), containing 30 items of 4–6 pitches, constructed without tonal or rhythmic contexts (durations are all equal); Standardised Tests of Musical Intelligence (Wing, 1961), tonal memory subsection, containing 30 items of 3–10 pitches, created in both tonal and rhythmic contexts; and the Norris Pitch Retention and Discrimination Test (Norris, 1997), containing 32 items of 4–7 pitches, with tonal but no rhythmic context. The fourth recognition test, Musical Aptitude Profile tonal imagery—melody section (Gordon, 1965), was selected on the basis of its different content; that is, the listener must determine whether two melodic phrases have similar or different musical material rather than pinpoint a precise difference in two tonal stimuli (40 items). The test manuals that accompany each of these tests address the questions of both validity and reliability. All recognition tests were administered in identical random order in the subjects' choral classrooms.

The Mowrer Tonal Memory Test (Mowrer, 1996), selected because it is a reproduction test that has been subjected to the constructs of

Table 1
Means and Standard Deviations of All Test Scores by Levels of Grade, Gender, and Musical Participation

Level	N	Test (Highest Possible Score)									
		Gordon (40)		Norris (32)		Seashore (30)		Wing (30)		Mowrer (35)	
		M	SD	M	SD	M	SD	M	SD	M	SD
Sample	210	29.1	5.2	13.1	4.6	18.6	6.1	16.0	4.2	18.3	6.5
Grade											
5	29	26.9	5.2	9.4	4.0	15.4	4.3	12.7	4.1	15.6	6.0
6	23	28.0	4.2	11.0	4.7	18.2	5.8	14.8	3.9	16.5	6.5
7	29	27.5	6.6	11.7	3.7	17.0	6.7	16.4	4.2	18.0	6.0
8	27	30.1	5.3	14.7	3.5	18.9	6.2	16.6	4.2	19.4	8.0
9	27	30.7	4.1	13.5	4.6	18.0	6.1	16.0	4.0	17.1	5.5
10	27	28.6	4.8	14.1	3.1	18.6	6.3	16.4	3.8	18.1	5.7
11	25	31.0	4.2	15.5	4.4	22.1	5.3	18.4	3.8	21.3	6.1
12	23	30.0	5.5	16.2	4.3	21.4	5.8	17.5	3.7	20.7	6.5
Gender											
Female	126	28.6	5.1	13.2	4.5	18.3	6.2	16.0	4.2	17.9	6.7
Male	84	29.9	5.4	13.1	4.6	19.1	6.0	16.1	4.3	18.9	6.1
Musical participation											
Band/ Chorus	142	29.8	5.0	13.7	4.5	19.6	6.0	16.8	4.1	19.6	6.6
Chorus only	68	27.8	5.5	12.0	4.6	16.5	5.8	14.4	4.2	15.4	5.2

reliability and validity, requires subjects to listen to, then replicate via the singing voice, short musical stimuli. The seven items, created in tonal and rhythmic contexts in the keys of C major and A minor, were transposed to the keys of G major and E minor to accommodate the changing male voice. Subjects were individually tested in a quiet, secluded room. Responses to the seven audiotaped acoustic piano items (presented to subjects on audiotape to eliminate potential human error in live playing) were audiotaped for later scoring. Scoring of the Mowrer test was based on a scale devised by Feierabend (1984) involving the assignment of point values (1–5) to subjects' responses to each of the seven test items. Points assigned and their descriptions are:

- 5—The tonal pattern is accurately reproduced with good intonation.
- 4—The tonal pattern is correctly reproduced, but with some uncertainty.

- 3—Melodic direction is evident, but some tones are incorrectly reproduced.
- 2—Melodic direction is evident, but no tones are correctly produced.
- 1—Reproduction of the tonal pattern is not recognizable.

I rated all student performances prior to two other raters to determine the effectiveness and conclusiveness of this scale. One type of response was common but not conclusively accounted for by the scale: response resembled stimulus pattern but was sung at a different pitch level. Because this type of performance most closely resembled "melodic direction is evident but no tones are correctly reproduced," it was assigned 2 points. Raters were trained using examples of each performance category. Using the total scores for each student given by each rater (seven items at 5 points each, for a sum of 35 points maximum), interrater reliability was established via coefficient alpha ($r = .97$) and by averaging correlations between the ratings of pairs of individual raters ($r = .91$). The raters' scores for each student were then averaged to obtain individual student scores on the Mowrer test.

Correlation and regression coefficients were determined with regard to the whole sample, singing experience (determined by a student's grade), musical participation (whether a student was in band and chorus or chorus only), and gender. Qualitative comments made by the instructors of the subjects are considered in the discussion of the reported data.

RESULTS

Basic descriptive statistics were computed to determine means and standard deviations for each test by levels of grade, gender, and level of musical participation (Table 1). Because of observed homogeneity of variance for each of the tests in the contexts of gender, musical participation, and singing experience, correlation and regression coefficients were computed.

For the entire sample, Pearson correlation coefficients between the Mowrer reproduction test and the four recognition measures indicated moderate to moderately strong relationships (Table 2). The Seashore and Wing tests shared substantially stronger relationships (.66 and .64) with the Mowrer test than the Norris (.59) or the Gordon (.47).

Further inspection of Table 2 shows the role of singing experience or grade level in the relationship between performances on the Mowrer test and the recognition tests. In Grade 5, the weakest relationships may be noted, with coefficients ranging from .21 to .37. In

Table 2
Correlation Coefficients between Tests of Recognition and the Mowrer Test of Tonal Memory by Levels of Grade, Gender, and Musical Participation

Variable	N	Correlated Tests			
		Gordon/ Mowrer	Norris/ Mowrer	Seashore/ Mowrer	Wing/ Mowrer
Sample-wide	210	.47	.59	.66	.64
Grade					
5	29	.21	.31	.29	.37
6	23	.05	.53	.56	.49
7	29	.61	.67	.72	.79
8	27	.65	.68	.79	.71
9	27	.43	.69	.62	.50
10	27	.25	.29	.67	.58
11	25	.71	.61	.71	.68
12	23	.58	.69	.65	.69
Gender					
Female	126	.41	.60	.67	.58
Male	84	.51	.59	.65	.68
Musical participation:					
Band/ Chorus	143	.47	.61	.65	.54
Chorus only	67	.41	.50	.61	.50

the case of the Gordon test, there is considerable inconsistency as noted in Grades 6, 9, and 10, with coefficients ranging from .05 to .71. The Norris test shares fairly consistent relationships with the Mowrer test across grade levels, with the exception of Grade 10, which had a relatively low coefficient of .29. The coefficients in Table 2 indicate that consistently moderate to moderately strong 7th–12th-grade relationships exist between the Seashore and Wing recognition measures.

In looking further at these relationships by levels of musical participation and gender, it can be seen that there are similar degrees of strength between the Mowrer test and the four recognition measures

Table 3

Summary of Stepwise Multiple Regression Analysis for Recognition Test Variables Predicting Performance on the Reproduction Tonal Memory Test (N = 210)

Variable	B	SE B	β^*
Step 1			
(Constant)	5.38	1.08	
Seashore Test	.70	.06	.66
Step 2			
(Constant)	3.13	1.11	
Seashore Test	.51	.06	.48
Norris Test	.43	.09	.31
Step 3			
(Constant)	1.00	1.24	
Seashore Test	.38	.07	.37
Norris Test	.32	.09	.23
Wing Test	.37	.11	.25

Note. $R = .66$, $R^2 = .44$ for Step 1; $R = .71$, $R^2 = .51$ for Step 2; $R = .73$, $R^2 = .53$ for Step 3.

* Represents the standardized coefficients.

(Table 2). On all measures, band and chorus students scored significantly higher than their chorus-only peers did [t -values ranging from 2.48 ($p = .017$) on the Wing test to 4.18 ($p < .001$) on the Mowrer]. Conversely, there were no significant differences in scores on any of the same tests as a result of gender [t -values ranging from .07 ($p = .91$) on the Wing test to 1.67 ($p = .10$) on the Gordon test]. Further analyses of grade level and the other variables (gender and musical participation) were not made due to the resultant small number of subjects in each cell, which would likely result in artifactual (potentially misleading) data.

To determine the predictive capability of any one of the recognition tests (or combination thereof), stepwise multiple regression was used (Table 3). The relationship between the Mowrer and the combination of the Seashore, Norris, and Wing yielded the greatest predictive relationship ($R = .73$, $R^2 = .53$). Correlating these three recognition tests with the Mowrer test (Table 4) resulted in moderate to moderately strong correlations (R ranging from .65 to .83) and regression coefficients (R^2 ranging from .42 to .69) for Grades 6 through 12.

Table 4
Multiple Correlation and Regression Coefficients for the Combination of Scores on Seashore, Norris, and Wing Tests and the Mowrer Reproduction Scores by Levels of Grade

	Grade								
	5	6	7	8	9	10	11	12	Total
<i>R</i>	.41	.65	.83	.83	.75	.66	.73	.76	.73
<i>R</i> ²	.17	.42	.69	.69	.56	.44	.53	.58	.53

DISCUSSION AND CONCLUSIONS

Although the lack of randomization is evident, it is important to note that great care was taken to control for the instruction that the students received, as evidenced in the tenure of the school music faculty, the ages of the subjects, and the cross-sectional design of this study. The results, therefore, may be deemed to some degree generalizable to other choral music programs, and worthy of discussion and further inquiry. Because few studies related to the present topic exist, suggested relationships between this study and others are hereby acknowledged in many cases as being very speculative, but necessary in setting up a frame of reference for future investigations.

Generally speaking, it can be concluded that sample-wide there is a moderate relationship between performance on a reproduction test and performances on certain recognition tests. The studies of Bentley (1969), Pedersen and Pedersen (1971), and Zwissler (1971) support this conclusion. The statistical strength of the relationships shown in Table 2 has long been accepted as satisfactory in the establishment of concurrent or predictive validity of prominent tests of musical aptitude (e.g., Gordon, 1967, 1969; Wing, 1954; Young, 1972). Therefore, it is possible that under conditions of rigorous control, use of a reproduction test such as the Mowrer may provide information about tonal skills of choristers.

Consideration of singing experience allows additional speculation about the validity of a reproduction tonal memory test such as the Mowrer. It can be seen in Table 2 that, with a few exceptions, the relationship between performances on the two types of tests is fairly stable across the range of singing experience studied. This consistency and the fact that there were no significant differences by levels or

gender on any test contradict previous research (i.e., Davies & Roberts, 1975) indicating that junior-high-age males frequently have problems with vocal pitch matching. In fact, the males in this sample scored, on average, 1 point higher than the females did on the Mowrer test (Table 1). A decrease in accurate singing by males at this age would have made an impact on the statistical relationship between the reproduction and recognition tests, in both the contexts of grade level and gender. Although the stimuli were pitched appropriately for each male subject, caution in the interpretation of the results is suggested. Other factors found to affect development of pitch-matching accuracy (the timbre of the stimulus, in particular) are not accounted for in this study.

Inconsistent with the overall picture are weaknesses (although in most cases, the coefficients are not unimpressive) in the relationship between the two types of tests at certain grade levels. This observation may be accounted for, in part, by considering commentary made by the music teachers about the ability levels of the various classes. First of all, the relationships between the test scores in the Grade 5 class are the weakest in the sample. The students in Grade 5 were described as typical in ability, but inexperience logically played an important role in the observed test relationships. The modest coefficients obtained are consistent with those obtained by Cooper (1992) in her study of relevant relationships in performances by elementary students.

Cited by the subjects' music instructors as exceptionally strong classes, both vocally and musically, were the 7th, 8th, and 11th grades. In these classes, the correlation coefficients were consistently moderate to moderately strong, regardless of the pairing of tests. The stronger relationships are consistent with those obtained by Geringer (1983) with high-ability 4th graders, and with Bentley (1969), whose research demonstrated that students who possessed higher pitch-discrimination skills were more accurate singers. Conversely, the 9th- and 10th-grade classes were described by their instructors as low-ability and low-achieving, both vocally and musically. The lower correlation coefficients for these grade levels (Table 2) are consistent with the speculation that ability or achievement has an impact on the relationship between sung tonal memory and the recognition tonal memory tasks (these results are also consistent with those of Geringer and Bentley).

The role of musical participation seems to be related to achievement and ability, as students who participated in both band and chorus not only scored significantly higher on all tests, but also the relationships between their scores on the recognition tests and the reproduction tonal memory test were somewhat stronger (Table 2). Relevant to the issue of musical participation is the fact that when

they enter Grade 7, subjects receive two times more instruction in both band and chorus than they previously received in Grades 5 and 6. Future research with a larger sample could determine not only the impact of ability, achievement, and amount of instruction on the validity of the reproduction tonal memory test, but also how such variables change across the singing experience levels observed in this study. Path analysis would be the most logical approach to obtaining more conclusive information.

In regard to the weaker and wider range of correlation coefficients related to the Gordon test (.05 to .71), it seems probable that this phenomenon is due to the nature of the Gordon test items not being as closely related to the reproduction tonal memory test items (as are the more precise tasks of the other three recognition tests of tonal memory). This, coupled with the variety in class abilities (Grade 6 was deemed as a class of very high abilities and very low abilities—there were no medium-ability students), may account for the diversity observed.

As the data stand, it appears that the Wing and Seashore tests offer a potential source of concurrent validity for reproduction tonal memory tests across levels of singing experience. In light of the possibility that reproduction tests may not be rigorously administered and scored, the regression data point to a potential source of evaluation that would be of great use and of higher validity than reproduction tests. The Seashore, Norris, and Wing tests in combination offer a viable alternative to reproduction tonal memory tests. The total sample correlation and regression coefficients ($R = .73$, $R^2 = .53$), although impressive, allow room for developing a measure that would have even greater predictive power.

Perhaps the keys to developing such a test lie in the lengths and rhythmic and tonal make-ups of the test items. The Seashore items are without tonal reference and shorter in length compared to the more musical, tonal, and longer lengths of the Norris and Wing items. Of additional concern is the inclusion or exclusion of rhythmic context. Future studies hold the potential for developing highly valid recognition tests and test items that yield information that has heretofore been determined by reproduction tonal memory tests. The encouraging, albeit tentative, findings of this study indicate the strong possibility of the development of such a test. Such endeavors will also insure that determination is made of the conditions under which such a measure is optimally valid. In addition to the possibility of accurately assessing musical potential in a choral ensemble, development of an instrument as described can potentially assist in the diagnosis and eventually lead to the amelioration of certain perceptual faults that often discourage young people from the pursuit of lifelong musical pleasures in the realm of singing.

REFERENCES

- Bartlett, D. L. (1996). Tonal and musical memory. In D. Hodges (Ed.), *Handbook of music psychology* (pp. 177–195). Lawrence, KS: National Association for Music Therapy.
- Bentley, A. (1969). Monotones. *Music Education Research Papers 1*. London: Novello and Company.
- Boisen, R. (1981). The effect of melodic context on students' aural perception of rhythm. *Journal of Research in Music Education*, 29, 165–172.
- Brinson, B. A. (1996). *Choral music methods and materials: Developing successful choral music programs, grades 5 to 12*. New York: Schirmer Books.
- Cooper, N. A. (1992). Selected factors related to children's singing accuracy. *Dissertation Abstracts International*, 53 (12A), 4244. (University Microfilms No. AAG93-07489)
- Davies, A. D. M., & Roberts, E. (1975). Poor pitch singing: A survey of its incidence in school children. *Psychology of Music*, 3 (2), 24–36.
- Feierabend, J. M. (1984). The effects of specific tonal pattern training on singing and aural discrimination abilities of first grade children. *Dissertation Abstracts International*, 45, 110A. (University Microfilms No. DA84-10243)
- Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education*, 31, 93–99.
- Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education*, no. 104, 17–37.
- Gordon, E. E. (1965). *Musical Aptitude Profile*. Boston: Houghton Mifflin.
- Gordon, E. E. (1967). *A three-year longitudinal predictive validity study of the Musical Aptitude Profile*. Iowa City: University of Iowa Press.
- Gordon, E. E. (1969). Intercorrelations among Musical Aptitude Profile and Seashore Measurements of Musical Talents subtests. *Journal of Research in Music Education*, 17, 263–271.
- Gordon, E. E. (1984). *Learning sequences in music: Skill, content, and patterns*. Chicago: G.I.A. Publications.
- Lamb, G. H. (1986). *Choral techniques*. Dubuque, IA: William C Brown Publishers.
- Long, P. A. (1977). Relationships between pitch memory in short melodies and selected factors. *Journal of Research in Music Education*, 25, 272–282.
- Mowrer, T. A. (1996). Tonal memory as an audition factor for choral ensembles (Doctoral dissertation, Temple University, 1996). *Dissertation Abstracts International*, 57 (05A), 1988.
- Norris, C. E. (1997). *Norris Pitch Retention and Discrimination Test*. Unpublished booklet.
- Pedersen, D. M., & Pedersen, N. O. (1971). The relationship between pitch recognition and vocal pitch production in sixth-grade students. *Journal of Research in Music Education*, 19, 265–272.
- Pembroke, R. G. (1986). Interference of the transcription process and other selected variables on perception and memory during melodic dictation. *Journal of Research in Music Education*, 34, 238–261.
- Price, H. E., Yarbrough, C., Jones, M., & Moore, R. S. (1994). Effects of male timbre and falsetto, and sine wave presentations on interval matching by inaccurate singers. *Journal of Research in Music Education*, 42, 269–284.
- Roe, P. F. (1983). *Choral music education*. Englewood Cliffs, NJ: Prentice Hall.

- Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly Journal of Music Teaching and Learning*, 1 (1 & 2), 81-95.
- Rutkowski, J. (1996). The effectiveness of individual/small-group singing activities on kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education*, 44, 353-368.
- Seashore, C. E., Lewis, L., & Saetveit, J. G. (1960). *Seashore measures of musical talents*. New York: The Psychological Corporation.
- Sink, P. E. (1983). Effects of rhythmic and melodic alterations on rhythmic perception. *Journal of Research in Music Education*, 31, 101-113.
- Taylor, J. A. (1976). Perception of tonality in short melodies. *Journal of Research in Music Education*, 24, 197-208.
- Wing, H. D. (1954). Some applications of test results to education in music. *British Journal of Psychology*, 24, 161-170.
- Wing, H. D. (1961). *Standardised tests of musical intelligence*. The Mere, England: National Foundation for Educational Research.
- Yarbrough, C., Bowers, J., & Benson, W. (1992). The effect of vibrato on the pitch matching accuracy of certain and uncertain singers. *Journal of Research in Music Education*, 10, 30-38.
- Yarbrough, C., Green, G., Benson, W., & Bowers, J. (1991). Inaccurate singers: An exploratory study of variables affecting pitch matching. *Bulletin of the Council for Research in Music Education*, no. 107, 23-34.
- Yarbrough, C., Morrison, S., Karrick, B., & Dunn, D. (1995). Effect of male falsetto on the pitch matching accuracy of boy singers, grades K-8. *Update: Applications of Research in Music Education*, 14 (1), 4-10.
- Young, W. T. (1972). A statistical comparison of two recent musical aptitude tests. *Psychology in the Schools*, 9 (2), 165-169.
- Zwissler, R. N. (1971). An investigation of the pitch discrimination skills of first grade children identified as accurate singers and those identified as inaccurate singers. *Dissertation Abstracts International*, 32, 4056A. (University Microfilms No. AAG72-02947)

Submitted December 14, 1998; accepted December 1, 1999.