

# Journal of Research in Music Education

<http://jrm.sagepub.com/>

---

## **An Application of Amabile's Consensual Assessment Technique for Rating the Creativity of Children's Musical Compositions**

Maud Hickey

*Journal of Research in Music Education* 2001 49: 234

DOI: 10.2307/3345709

The online version of this article can be found at:

<http://jrm.sagepub.com/content/49/3/234>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



National Association  
for Music Education

[National Association for Music Education](#)

Additional services and information for *Journal of Research in Music Education* can be found at:

**Email Alerts:** <http://jrm.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jrm.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jrm.sagepub.com/content/49/3/234.refs.html>

>> [Version of Record](#) - Jan 1, 2001

[What is This?](#)

*Amabile has proposed that the most valid way to measure creativity is by using experts' subjective assessment of creative products—a technique she has labeled "consensual assessment." The purpose of this study was to test the reliability of the consensual assessment technique on children's musical compositions and to determine which group of judges provides the most reliable ratings of creativity for music compositions of fourth- and fifth-grade children. The groups that were compared were music teachers, composers, theorists, seventh-grade children, and second-grade children. The interjudge reliabilities for each group's creativity ratings on 11 children's musical compositions were: composers, .04; all music teachers, .64; music theorists, .73; seventh-grade children, .61; and second-grade children, .50. Significant correlations were found between the music teachers and music theorists, and between the two groups of children. There were very weak or negative correlations between the composers' scores and the scores of the other groups.*

**Maud Hickey, Northwestern University**

## An Application of Amabile's Consensual Assessment Technique for Rating the Creativity of Children's Musical Compositions

Since the 1960s, the measurement of creativity has been heavily influenced by Guilford's hypothesis that creative thinking consists of divergent thinking factors, which exist along with several other factors posited in his Structure of Intellect Model (1967). Guilford's creativity test (Guilford, Merrifield, & Wilson, 1958) and Torrance's widely used standardized Tests of Creative Thinking, or TTCT (1974, 1981), are based on Guilford's hypothesis and measure the divergent thinking factors of fluency, flexibility, elaboration, and originality.

---

Maud Hickey is an assistant professor of music education and technology in the School of Music, Northwestern University, 711 Elgin Road, Evanston, IL 60208; e-mail: mhhickey@northwestern.edu. Copyright © 2001 by MENC—The National Association for Music Education.

The TTCT are paper-and-pencil tests provided in both verbal ("Thinking Creatively with Words") and figural ("Thinking Creatively with Pictures") versions. In each version, subjects are asked to guess causes or consequences, provide ideas for product improvement, or suggest unusual uses for a variety of prompts. For example, one activity asks test-takers to list as many interesting and unusual uses that they can think of for a cardboard box. Responses are scored for fluency, flexibility, originality, and elaboration. Fluency is determined by the sheer quantity of responses; flexibility, by the number of different categories of responses; originality, by the uniqueness of the response (relative to a set of responses); and elaboration, by the details given in the response.

The quantitative assessment of children's musical creativity began in the 1970s and was shaped by Guilford's work (Gorder, 1976; Vaughan, 1971; Webster, 1977). Since then, the influence of Guilford and Torrance is apparent in most studies of music creativity in which some or all aspects of the divergent factors are used to score musical responses in order to assess musical creativity (Baltzer, 1990; Holliger, 1989; Josuweit, 1992; Moore, 1990; Schmidt & Sinor, 1986; Vaughan, 1971, 1977; Vold, 1986; Webster, 1977, 1987, 1990, 1994; Webster, Yale & Haefner, 1988). Currently, Webster's 1994 Measurement of Creative Thinking in Music-II (MCTM-II) is the most well-known measure of creative musical potential and, like the TTCT, measures divergent thinking factors, as well as the convergent factor of musical syntax.

The validity of this factorial approach to the measurement of creative thinking has come under recent criticism (Amabile, 1996; Brown, 1989; Perkins, 1988; Plucker & Renzulli, 1999; Plucker & Runco, 1998). The greatest criticism is that the theoretical constructs for the divergent thinking measures were validated through specialized tests such as factor analyses, but never against any external measure of creative productivity. "The basic problem seems to be that creativity tests had only apparent construct validity and certainly not criterion validity" (Brown, 1989, p. 8). In addition, these paper-and-pencil divergent thinking tests do not capture the greater and more complex instances of real-life creative endeavors.

As an alternative to using divergent thinking measures for creative thinking, Amabile proposed that the most valid way to measure creativity is by using experts' global and subjective assessment of creative products—a technique she has labeled "consensual assessment" (1983, 1996). This technique requires judges to rate the creativity of an artistic product by using their own subjective definition rather than any given objective criteria or checklist. A consensual assessment item for rating the creativity of a painting might read: "On a scale of 1 to 5, and using your own subjective definition of creativity, rate the degree to which the painting is creative." Amabile argues that, ultimately, it is impossible to articulate clear, objective criteria for a creative product; rather, "a product or response is creative to the extent that appropriate observers independently agree it is cre-

ative" (1983, p. 31). Furthermore, Amabile theorizes that the interjudge reliability of consensual assessment by experts is the equivalent of construct validity: "If appropriate judges independently agree that a given product is highly creative, then it can and must be accepted as such" (1982, p. 1002). In a comprehensive review of studies using consensual assessment for rating the creativity of a variety of artistic, verbal, and problem-solving tasks as assessed by different groups of judges and among varied populations of subjects, reliability figures were consistently high (Amabile, 1996).

The use of consensual assessment for measuring musical creativity is recent and limited. Bangs (1992) devised a Dimension of Judgment tool that consisted of 19 subjective items. Interjudge reliability among the three judges who rated children's musical compositions was .76 and .82 (two separate trials). Brinkman (1999) used a 3-item (originality, craftsmanship, and aesthetic value) form of consensual assessment technique to rate 32 high school instrumental students' melodies. The reliability of the composite "creativity" score for the three judges ranged from .77 to .96. The reliability of "creativity" ratings for children's musical compositions ranged from .62 to .73 in a 1995 study by Hickey, and was .93 in another (Hickey, 1996). The consensual assessment technique has been modified and also used successfully for rating the creativity of musical compositions by Daignault (1997) and for rating musical improvisations by Amchin (1996) and Priest (1997).

Ratings of children's musical compositions using a consensual assessment technique have been compared to children's musical creativity ratings from Webster's MCTM-II. Hickey (1995) found no significant correlations between consensual assessment creativity ratings of children's musical compositions and their composite or subscores of fluency, extensiveness, originality, and syntax on Webster's MCTM-II (1994). In a comparative analysis of several types of musical product rating procedures, Hickey and Webster (1995) found that scores from implicit (subjective) rating forms proved to be most predictive for the constructs of originality/creativity qualities and aesthetic value of children's compositions and that scores from more explicit forms were most predictive for the constructs of craftsmanship and technical quality.

Although subjective and consensual assessment by "experts" is an appealing strategy for rating the creativity of musical products, several questions arise: When using the consensual assessment technique, who are the "experts" or "appropriate judges" to rate the creativity of products? Are the professional creators (i.e., composers, painters, and sculptors) the experts? Are teachers the experts? And can children reliably assess the creativity of other children's creative products?

Runco, McCarthy, and Svenson (1994) sought to determine which group of judges were most reliable for judging the creativity of visual artwork using the consensual assessment technique. Forty-seven college-level subjects each completed three artworks to be self-rated,

rated by peers, and rated by professional artists for creativity. The self-assessment rankings and peer-assessments rankings for subjects' art works were similar, and professional judges' rankings were also similar, but the differences between rankings were not significant, and the scores given by the professionals were much lower than those given by the students. Runco et al. (1994) concluded that perhaps the professionals were not as sensitive to differences among products and were overly critical.

In no study have researchers explicitly examined the differences between or the relationships among different groups of judges when using the consensual assessment technique to rate music compositions. The purposes of this study were to examine these issues, while testing the reliability of the consensual assessment technique for rating children's musical compositions. Specifically, the researcher sought to determine: (a) the interjudge reliability of the consensual assessment technique when used on children's musical compositions; (b) which group of judges—composers, theorists, teachers, or children—would make the most reliable creativity ratings of fourth- and fifth-grade children's musical compositions; and (c) the relationships of mean creativity scores between different groups of judges when using the consensual assessment technique to rate children's musical compositions.

## METHOD

The 12 compositions used for assessment in this study were randomly selected from a total pool of 21 original compositions generated by fourth- and fifth-grade subjects in a previous study (Hickey, 1995). In the 1995 study, the subjects were given unlimited time to create an original composition using a synthesizer connected via MIDI interface to a Macintosh computer. The final compositions were captured in MIDI file format using a computer program that allowed the recording of up to three simultaneous tracks of music. No compositional parameters were given, and students were encouraged to rerecord their compositions as often as necessary until they were satisfied with their finished product.

The five groups of judges used to rate the compositions were music teachers, composers, theorists, seventh-grade children, and second-grade children. The 17 music teachers were graduate students enrolled in a summer music education master's degree course program. The teachers were divided into the following groups for analysis: 10 "instrumental" music teachers whose teaching experience ranged from 2 to 26 years (with an average of 7 years) of teaching instrumental music only; 4 "mixed-experience" teachers—teachers who taught a combination of instrumental and choral or instrumental and general music—whose teaching experience ranged from 3 to 10 years (average, 7 years); and 3 "general/choral" music teachers—elementary general music teaching with some choral music—whose experience ranged from 2 to 7 years (average, 4.33 years). The

three professional composers had at least 15 years of experience with writing music in a wide variety of genres ranging from jazz to classical. Two of the composers were college teachers, and one composer was a graduate student. The four music theorists were college theory professors with at least 10 years' experience in teaching music theory at this level. The two groups of children came from contained classrooms in a private grade school in the northeast United States. The seventh-grade group consisted of 7 girls and 7 boys, while the second-grade group contained 13 boys and 11 girls.

### Procedure

The music compositions were recorded in random order onto an audiocassette tape. The order of the compositions on each tape was different for each group of judges. The composers and the theory teachers independently listened to and rated the musical compositions using the consensual assessment forms. The music teachers listened to the compositions together but rated them independently. The groups of children listened to the compositions together in their respective classrooms. Before listening to and rating the compositions, the researcher engaged the children in discussion about "liking" music and/or thinking that music is "creative." The children shared ideas about what "creative" meant to them, and the discussion was guided to help them focus on understanding this term for rating music. They then independently rated each composition, first for "Liking," and on a second listening, for "Creativity."

All the judges were informed that the compositions were composed by fourth- and fifth-grade children. Following Amabile's suggestion for proper consensual assessment technique procedures (1996), the judges were instructed to rate products relative to one another rather than against some "absolute" standard in the domain of music.

The judges rated the compositions for items in addition to creativity; however, only the creativity item was used for comparison. Some craftsmanship and aesthetic-quality items were also included on the music composition rating forms, as recommended by Amabile (1983), in order to assure discriminant validity between these areas and creativity. More specifically, the theorists and composers used an 18-item consensual assessment rating form to rate the compositions. The form was developed by combining and adapting items from Amabile's *Dimensions of Creative Judgment* (1982) and Bangs' *Dimension of Judgment* (1992). This form was used and tested in two previous studies (Hickey, 1995; 1996). The 18 items fell under one of three dimensions: creativity, craftsmanship, and aesthetic appeal. The items consisted of 7-point Likert-type scales with anchors marked "low," "medium," and "high." The music teachers used a 3-item form with 7-point rating scales for creativity, craftsmanship, and aesthetic appeal. For all judges, the creativity item was worded in this way: "Using your own subjective definition of creativity, rate the

Table 1  
*Interjudge Reliability (Intraclass Correlation) Adjusted for Three Judges*

	Reliability coefficient using all judges	Reliability coefficient adjusted for three judges
Composers ( $N = 3$ )	.04	.04
Theorists ( $N = 4$ )	.78	.73
All teachers ( $N = 17$ )	.91	.64
Instrumental teachers ( $N = 10$ )	.86	.65
Mixed teachers ( $N = 4$ )	.60	.53
General/choral teachers ( $N = 3$ )	.81	.81
7th-grade Children ( $N = 13$ )	.87	.61
2nd-grade Children ( $N = 24$ )	.89	.50

degree to which the composition is creative.” The children rated the compositions for “Liking” and for “Creativity,” using a separate form for each scale. The Creativity form asked the students to rate each composition on a 5-item scale with “Not Creative” and “Very Creative” marked on the low and high ends. The second-grade children’s form had icons (from plain to more elaborate/silly faces) at each point on the scale to aid them in understanding the continuum.

## RESULTS

Interjudge reliabilities for the creativity item were calculated using “Hoyt’s analysis,” an intraclass correlation technique that renders a coefficient alpha (Nunnally & Bernstein, 1994). The interjudge reliability, using mean creativity ratings for all groups, was .48; the interjudge reliability for all groups minus the composers was .78 (this reliability coefficient was calculated because the composers, as a group, showed no relationship in their ratings). Because each group had a different number of judges and a higher number of judges will yield a higher reliability rating than a smaller number, the interjudge reliability rating for each group was calculated not only for all of the judges, but also for  $N = 3$  judges. The resulting interjudge reliabilities for each group’s creativity ratings on the musical compositions ranged from a low of .04 for the music composers to .81 for the general/choral music teachers. These reliability figures are listed in Table 1.

The correlations of mean creativity ratings among the different groups of judges are presented in Table 2. Due to the lack of agreement among the composers, each composer is represented separately rather than using the group mean for correlation with the other groups. Significant correlations were found between the groups of



Table 2

*Correlations of Mean Creativity Ratings between Groups of Judges*

Judges	1	2	3	4	5	6	7	8	9
1. Composer A									
2. Composer B	-.02								
3. Composer C	.07	-.26							
4. Music theorists	.16	-.02	.58						
5. All music teachers	.35	.01	.37	.90**					
6. Instrumental teachers	.45	-.09	.39	.88**	—				
7. Mixed teachers	.18	.11	.35	.86**	-.78**				
8. General/ choral teachers	.14	.17	.19	.63*	-.68*	.72*			
9. 7th-grade children	.09	.08	.37	.26	.03	-.01	.27	-.21	
10. 2nd-grade children	.19	-.03	.19	.38	.18	.11	.41	-.01	.83**

\*\*  $p < .01$ , \*  $p < .05$ .

music teachers, between the music teachers and music theorists, and between the two groups of children. Although the music teachers and music theorists agreed with each other, and the groups of children had a high positive correlation with each other, the theorists and teachers showed moderate to low correlations with the groups of children. Worth noting is not only the lack of any strong positive correlation among the composers and between the composers and the other groups, but also the several instances of negative correlations between the composers with themselves and others.

## DISCUSSION AND IMPLICATIONS FOR MUSIC EDUCATION

The purposes of this study were to test the reliability of the consensual assessment technique and to determine who might be the most reliable group of experts to judge the creativity of children's musical compositions when using a consensual assessment technique. It is clear from these findings that while the consensual assessment technique is indeed a moderately reliable technique for measuring the creativity of children's compositions by most groups of judges, composers were the least consistent group to do so. The groups of music teachers, music theorists, and seventh-grade children showed agreement within their respective groups, with the general/choral teachers clearly the highest, at .81. The group of second-



grade children and the “mixed” teachers were among the lowest, with .50 and .53 reliability coefficients, respectively. It seems that the best “experts,” or at least the most reliable judges, may be the very music teachers who teach the children—the general/choral music teachers. Perhaps the extensive music training that music teachers have, along with their experience in the classroom with children, provides them with the tools necessary to make consistent and valid judgments about the creative quality of children’s original musical products.

It is most interesting that the composers used in this study were the group least able to come to an agreement on the creativity of the children’s compositions. In music education, composition is sometimes viewed as mysterious, and often the only experts considered in this realm are the professional composers. Perhaps music teachers should have reason to feel more confident in their ability to accurately assess the relative creativity of their students’ musical compositions. The world in which professional composers work may be too far removed from the world of children’s musical creative thinking. It may be simple exposure to children’s compositions that is needed. Perhaps the composers could be “trained” to be as reliable judges for children’s compositions as the teachers if given more experience with children or more examples of what to expect from children as composers. However, if the lack of exposure to, or expectations of, children’s compositions were the confounding variable, one would hypothesize that the music theorists would also have difficulty in coming to agreement. This was not true in the present study. Future research should continue to look at the differences within and between these different groups, as well as at how training of expectations might affect the outcome of judging.

Although the two groups of children showed strong agreement between them, they did not correlate highly with the teachers. A closer examination of the relationship between the “Liking” and “Creativity” scores revealed that the children were unable to separate these two concepts. The compositions that they selected highest for “Liking” were also chosen as the most creative. This is not unusual, and, as Gardner (1982) points out, children desire to conform or be more “literal” in their creative artistic development beginning around the age of 9. They value more “likable,” as opposed to, perhaps, more “creative” and “unusual” compositions at this age level. Further research is warranted to see if this trend is different in younger children and older children, which, if true, might coincide with a hypothesis of a U-shaped developmental curve of creative thinking. A U-shaped developmental curve has been suggested in creativity development in which creativity is high in early childhood (marked by play and freedom from conformity), is followed by a slump around the age of 9 years, and re-emerges in a more sophisticated form for some in adulthood (Albert, 1996; Keegan, 1996; Runco & Charles, 1997). Furthermore, explicit understanding of what children consider to be creative will lead to better teaching for creative music composition. Teachers may need to teach creative

techniques explicitly when teaching musical composition in order to nurture and support not only well-crafted compositions, but creative ideas as well. An extension of this research should include interviews with creative child composers, as well as the children who judge, in order to inform teachers and researchers more about children's understanding and capabilities in creative music making.

The consensual assessment technique for rating the "creativity" of musical compositions proves to be reliable when used by the most knowledgeable (in that context) group of judges. And as Amabile (1996) has repeatedly pointed out, this reliability, or agreement among "expert" judges, supports the construct validity for such a technique. In music, researchers have developed suitable and reliable means for evaluating performances. Rating forms for contests, which list specific musical criteria that are objectively scored for a final award, serve their evaluative and educational purposes. However, when music teachers may be required to assess students' ability to compose or improvise music, there is confusion. In these situations, the subjectivity of the art form and the difference between contexts in which the music is created cause difficulty in devising clear and objective criteria for assessment. This problem is true in research as well. The apparent variance between factorial measures of creative thinking, consensual assessment, and holistic measures warrants continued research to discover the meaning behind these differences. The continued development and refinement of a consistently reliable technique for rating the creativity of children's composition will prove useful for research efforts, as well as for educational purposes.

Although a main concern of this study was to test the reliability of consensual assessment for measuring the creativity of compositions, the results can be applied to music classrooms. If researchers are able to come up with a reliable form for rating musical compositions, then the next step is to collect and examine those compositions from children that are consistently rated as highly "creative." What are the features of these successful compositions? Learning more from studying the most creative compositions, researchers may be able to formulate essential rubrics to aid in assessing children's musical compositions in schools. Furthermore, music compositions rated highly "creative" or successful could also be used as models for elementary music classrooms—something that is desperately needed for teachers who strive to do more music composition activities in their classrooms.

## REFERENCES

- Albert, R. S. (1996). Some reasons why childhood creativity often fails to make it past puberty into the real world. In M. A. Runco (ed.), *Creativity from childhood through adulthood: The developmental issues* (pp. 43–56). San Francisco: Jossey-Bass Publishers.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43 (5), 997–1013.

- Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.
- Amabile, T. M. (1996). *Creativity in context*. Boulder, CO: Westview Press.
- Amchin, R. A. (1996). Creative musical response: The effects of teacher-student interaction on the improvisation abilities of fourth- and fifth-grade students. (Doctoral dissertation, University of Michigan, 1995). *Dissertation Abstracts International*, 56/08, 3044A.
- Baltzer, S. (1990). *A factor analytic study of musical creativity in children in the primary grades*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- Bangs, R. L. (1992). *An application of Amabile's model of creativity to music instruction: A comparison of motivational strategies*. Unpublished doctoral dissertation, University of Miami, Coral Gables, Florida.
- Brinkman, D. (1999). Problem finding, creativity style and musical compositions of high school students. *Journal of Creative Behavior*, 33 (1), 62-68.
- Brown, R. T. (1989). Creativity. What are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 3-32). New York: Plenum Press.
- Daignault, L. (1997). Children's creative musical thinking within the context of a computer-supported improvisational approach to composition. (Doctoral dissertation, Northwestern University, 1996). *Dissertation Abstracts International*, 57, 4681A.
- Gardner, H. (1982). *Art, mind, and brain. A cognitive approach to creativity*. New York: Basic Books, Inc.
- Gorder, W. D. (1976). An investigation of divergent production abilities as constructs of musical creativity. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1976). *Dissertation Abstracts International*, 37 (1), 177A.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J. P., Merrifield, P. R., & Wilson, R. C. (1958) *Unusual Uses Test*. Orange, CA: Sheridan Psychological Services.
- Hickey, M. (1995). *Qualitative and quantitative relationships between children's creative musical thinking processes and products*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- Hickey, M. (1996, December). *Consensual assessment of children's musical compositions*. Unpublished paper presented at the Research Poster Presentation, New York State School Music Association Convention.
- Hickey, M., & Webster, P. (1995). Rating scales and their use in assessing children's compositions. *The Quarterly Journal of Music Teaching and Learning*, 6 (4), 28-44.
- Holliger, Y. (1989). An investigative study on developing divergent thinking responses in children using a cognitive approach in music education. (Unpublished doctoral dissertation, Columbia University Teachers College, 1987). *Dissertation Abstracts International*, 49 (9), 2574A.
- Josuweit, D. (1992). *The effects of an audiation based instrumental music curriculum upon beginning band students' achievement in music creativity*. (Unpublished doctoral dissertation, Temple University, May, 1991).
- Keegan, R. T. (1996). Creativity from childhood to adulthood: A difference of degree and not of kind. In M. A. Runco (Ed.), *Creativity from childhood through adulthood: The developmental issues* (pp. 57-66). San Francisco: Jossey-Bass Publishers.
- Moore, B. R. (1990). The relationship between curriculum and learner: music composition and learning style. *Journal of Research in Music Education*, 38, 24-38.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

- Perkins, D. N. (1988). The possibility of invention. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 362–385). Cambridge, UK: Cambridge University Press.
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (ed.), *Handbook of creativity* (pp. 35–61). Cambridge, UK: Cambridge University Press.
- Plucker, J. A. & Runco, M. A. (1998). The death of creativity measurement has been greatly exaggerated: Current issues, recent advances, and future directions in creativity assessment. *Roeper Review*, 21 (1), p. 36–39.
- Priest, T. L. (1997). *Fostering creative and critical thinking in a beginning instrumental music class*. Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.
- Runco, M. A., McCarthy, K. A., & Svenson, E. (1994). Judgments of the creativity of artwork from students and professional artists. *The Journal of Psychology*, 128 (1), 23–31.
- Runco, M. A., & Charles, R. E. (1997). Developmental trends in creative performance. In M. A. Runco (Ed.), *The creativity research handbook. Volume I* (pp. 115–152). Creskill, NJ: Hampton Press.
- Schmidt, C., & Sinor, J. (1986). An investigation of the relationships among music audiation, musical creativity, and cognitive style. *Journal of Research in Music Education*, 34, 160–172.
- Torrance, E. P. (1974). *The Torrance tests of creative thinking: Technical-norms manual*. Bensenville, IL: Scholastic Testing Services.
- Torrance, E. P. (1981). *Thinking creatively in action and movement: Administration, scoring, testing manual*. Bensenville, IL: Scholastic Testing Service, Inc.
- Vaughan, M. (1971). Music as model and metaphor in the cultivation and measurement of creative behavior in children (doctoral dissertation, University of Georgia). *Dissertation Abstracts International*, 32 (10), 5833A.
- Vaughan, M. (1977). Musical creativity: Its cultivation and measurement. *Bulletin of the Council for Research in Music Education*, no. 50, 72–77.
- Vold, J. N. (1986). *A study of musical problem solving behavior in kindergarten children and a comparison with other aspects of creative behavior* (Unpublished doctoral dissertation, University of Alabama).
- Webster, P. (1977). *A factor of intellect approach to creative thinking in music*. Unpublished doctoral dissertation, Eastman School of Music, University of Rochester, Rochester, NY.
- Webster, P. (1987). Refinement of a measure of creative thinking in music. In C. Madsen and C. Prickett (Eds.), *Applications of research in music behavior* (pp. 257–271). Tuscaloosa: University of Alabama Press.
- Webster, P. (1990, March). *Study of internal reliability for the Measure of Creative Thinking in Music (MCTM)*. Paper presented at the General Poster Session of the National Biennial In-Service Conference of Music Educators National Conference, Washington, DC.
- Webster, P. (1994). *Measure of Creative Thinking in Music-II (MCTM-II). Administrative Guidelines*. Unpublished manuscript, Northwestern University, Evanston, IL.
- Webster, P., Yale, C., & Haefner, M. (1988, April). *Test-Retest reliability of Measures of Creative Thinking in Music for children with formal music training*. Paper presented at the poster session, 1988 MENC National In-Service Meeting, Indianapolis, IN.

Submitted February 16, 2001; accepted June 21, 2001.