

2021

Rating Scales and Their Use in Assessing Children's Music Compositions

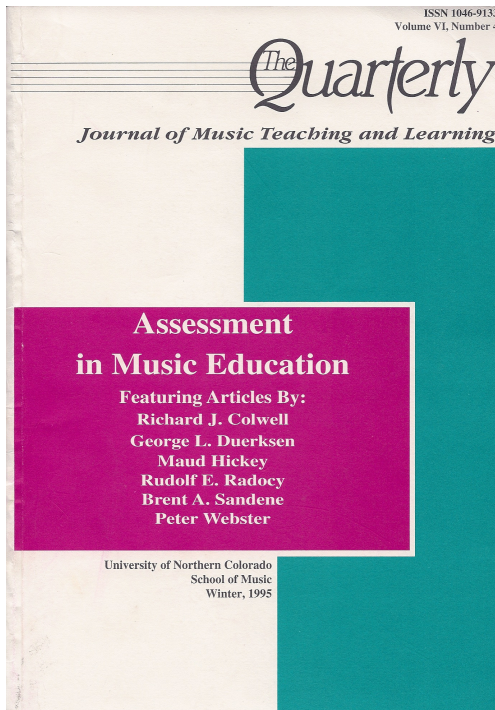
Peter Webster
Northwestern University

Maud Hickey
Ithaca College

Follow this and additional works at: <https://opencommons.uconn.edu/vrme>

Recommended Citation

Webster, Peter and Hickey, Maud (2021) "Rating Scales and Their Use in Assessing Children's Music Compositions," *Visions of Research in Music Education*: Vol. 16 , Article 26.
Available at: <https://opencommons.uconn.edu/vrme/vol16/iss6/26>



Title: Rating Scales and Their Use in Assessing Children's Music Compositions

Author(s): Peter Webster and Maud Hickey

Source: Webster, P., & Hickey, M. (1995, Winter). Rating scales and their use in assessing children's music compositions. *The Quarterly*, 6(4), pp. 28-44. (Reprinted with permission in *Visions of Research in Music Education*, 16(6), Autumn, 2010). Retrieved from <http://www-usr.rider.edu/~vrme/>

Visions of Research in Music Education is a fully refereed critical journal appearing exclusively on the Internet. Its publication is offered as a public service to the profession by the New Jersey Music Educators Association, the state affiliate of MENC: The National Association for Music Education. The publication of VRME is made possible through the facilities of Westminster Choir College of Rider University Princeton, New Jersey. Frank Abrahams is the senior editor. Jason D. Vodicka is editor of the *Quarterly* historical reprint series. Chad Keilman is the production coordinator. The *Quarterly Journal of Music Teaching and Learning* is reprinted with permission of Richard Colwell, who was senior consulting editor of the original series.

Rating Scales And Their Use In Assessing Children's Music Compositions

By Peter Webster and Maud Hickey

*Northwestern University
Ithaca College*

Interest in children's compositions and their evaluation has grown in the last twenty years as music educators have come to value more highly those teaching strategies that encourage creative thinking in music. Most will agree that music ability can no longer be viewed only as scores on standardized aptitude or achievement tests, points earned on an instrumental or vocal performance checklist, or answers to listening quizzes in a general music class. Important as this evidence is, it represents data from atomistic behaviors that are only part of the landscape of mental operations necessary to achieve in music. The assessment of achievement that comes from more holistic, authentic tasks that include more generative thinking (Boardman, 1989; Reimer, 1989; Webster & Richardson, 1992) is now seen as vital if we are to honestly confront the evaluation of music ability.

Research dealing with divergent, generative thinking by children has been influenced by a number of professional developments. Webster (1992, p. 266) has suggested at least three important developments in music education and psychology that are relevant in this context:

- the need to understand the actual cognitive processes of children, particularly those that are more generative in nature;
- the rise of more naturalistic studies of children's music making, including those that focus on the process of creative thinking in music; and
- a new emphasis on broadening the assessment of music achievement, especially those efforts to create and evaluate portfolios of student work.

Objective versus Subjective Techniques

At the heart of each of these developments are difficult decisions that each teacher and researcher must make about assessment. Decisions on what approaches to use are based in large part on whether the focus of the work is on the objective analysis of content in the products of composition or if the goal is to make some kind of overall quality judgment about the product.

For example, it is certainly possible to identify musical characteristics such as the number of notes used; length of composition; uses of silence, development, repetition and contrasts; and the use of timbre, harmony, or metric organization. These variables may reveal a number of important aspects about compositional thought, and be very useful in teaching and research. In fact, researchers have frequently used such variables in the study of both products and processes of children's composition (Kratus 1989, 1994).

Peter Webster is Professor of Music Education at the School of Music, Northwestern University. Maud Hickey is Assistant Professor of Music Education at the School of Music, Ithaca College. Their research interests include creative thinking in music and assessment.

Such objective, content analyses are useful and should be continued, but such data are not enough. Continued attention to the development of reliable, subjective assessments is necessary, especially if our work is to be seen as authentic. If we only describe objectively, we defeat the whole purpose of asking children to show evidence of high-level thinking. By asking children to compose and then assess the quality of these compositions, Amabile (1983) reminds us that "... not only does the task itself mimic real-world performance, the assessment technique mimics real-world evaluations of creative work" (p. 59).

Overall Purpose

For us, the difficulty in all of this is not in deciding whether subjective judgments are worth studying as legitimate, but rather just how they should be used. As we reviewed the literature on assessment of children's musical compositions, we discovered a wide scope and variety of techniques used to evaluate products. The rating scale emerged as one type of product evaluation tool which was of particular interest and seemed pervasive in the methodologies.

As we reviewed this literature, we were struck by both the similarities and differences of design and content of the ratings scales used. In terms of style, these scales contained items which were designed as either open-ended or very focused. The open-ended items seemed to rely on the implicit understanding that the judge might have for the construct under investigation, whereas the more specific items made more *explicit* (often in great detail) what the judge should consider.

In terms of music content, the scales we reviewed addressed both global issues such as music syntax and elaboration as well as more specific music content such as the manipulation of music elements (e.g. rhythm,

harmony). The scales that seemed to be more specific in nature were quite close to the more objective assessment items described above. The researchers in this case seem to be less interested in noting simply the existence of a rhythmic or harmonic quality, but more willing for a judge to rate the extent of its presence.

We were surprised to see that little or no research existed in music education that studied the effectiveness or quality of these different rating scales. We found few studies that investigated what we know about the interjudge reliability of explicit vs. implicit scales or whether there are differences when considering the content of these scales (specific vs. more global).

Perhaps most importantly, there is little evidence that helps us understand these ratings scales in terms of the overriding constructs of craftsmanship (technical skill), originality/creativity (imaginativeness) and aesthetic value (feelingful musical experience)

of the children's compositions. These three overriding constructs are often cited as the critical elements in judging creative thinking in the social psychology of creativity and in music and art (Amabile, 1982, 1983; Getzels & Csikszentmihalyi, 1976; Webster, 1987). What might be the relationship between compositions judged as the *best* and *worst* in these three constructs and the kinds of rating scales described above?

Convinced of the importance rating scales play in the valid assessment of musical quality in children's compositions, we decided to create a set of scales that was based on the two kinds of style (explicit and implicit) and the two types of music content (specific and global) and submit these scales to more rigorous study. Our overall intent was to provide better information about interjudge reliability and concurrent validity when com-

The assessment of achievement that comes from more holistic, authentic tasks that include more generative thinking ... is now seen as vital if we are to honestly confront the evaluation of music ability.

By asking children to compose and then assess the quality of these compositions, Amabile (1983) reminds us that “. . . not only does the task itself mimic real-world performance, the assessment technique mimics real-world evaluations of creative work” (p. 59).

pared with the overriding constructs of craftsmanship, originality/creativity, and overall aesthetic value.

Past Research

Study of Children's Compositional Processes and Products

We reviewed past research which has systematically measured and/or evaluated children's musical composition. A wide variety of assessment tools emerged. These studies can be organized into categories of methodology (positivistic and naturalistic) and content focus (product, process or both).

Naturalistic studies varied methodologically from ethnographic observations of free improvisation/composition behaviors (Christensen, 1992; Davies, 1992; Levi, 1991) to observations of more teacher/researcher devised tasks (Cohen, 1980; DeLorenzo, 1989; Loane, 1984; Swanwick & Tillman, 1986; Wiggins, 1992). Descriptions of compositional processes emerged in the form of models and categorizations (Christensen, 1992; Cohen, 1980; DeLorenzo, 1989; Levi, 1992; Wiggins, 1992), while the description of compositional products included musical content analyses (Davies, 1992; Loane, 1984) and developmental trends (Swanwick & Tillman, 1986). The findings provided rich descriptions of children's compositional processes and products and offered keys to the development of both objective and subjective assessment techniques. Surprisingly, little attempt was made to make actual qualitative evaluations in the form of summative statements about children's final compositions.

Positivistic assessment of children's musical compositions also varied in scope of methodology as well as in purpose. As noted above, researchers have provided content analyses of children's compositions and have

recently begun to analyze quantitatively the compositional processes of children (Hickey, 1993; Hoffman, Hedden, & Mims, 1990, 1991; Kratus, 1994, 1989, 1991) as well as the relationship of compositional processes to the quality of the final products (Hedden, 1992; Hickey, 1995; Kratus, 1991, 1994).

Within this group of studies designed to assess children's musical creative thinking and compositions, researchers utilized either previously constructed tools such as those by Webster (1989) and Wang (1985), techniques such as process analysis from Kratus (1989), or designed their own measures. The use of rating scales to assess the quality of compositions emerged as measurement tools in eight separate studies: Webster, 1977, 1989; Hassler & Feil, 1986; Moore, 1990; Kratus, 1991, 1994; Bangs, 1992; and Smith, 1993). It is on these eight studies that we now focus.

Studies Using Rating Scales

Webster (1977) designed *Thinking Creatively with Music* to evaluate the compositions, improvisations, and analytical efforts of 77 high school-aged students. Taken from this measure for the present investigation were items which contained rating scales for the judgment of originality and elaboration. Both items contained explanations of the criteria to be judged as well as descriptions for each point on the scale. Webster's *Measurement of Creative Thinking in Music-II* (1989), designed for primary grade children, contained items used to rate *Free Composition*. After completing several previous tasks using a piano, sponge ball, and temple blocks, subjects are asked to make up a song without any parameters except that it have a beginning, middle and end. *Free composition* is rated for musical syntax on one item, and musical originality on another. Both items

contained rating scales with a list of criteria appropriate to the rating of that item.

Hassler and Feil (1986) used open-ended rating scales to measure the *creative musical ability* of 30 high school subjects. The subjects were asked to present an original, notated composition to four judges who scored the tape-recorded performances on first impression, originality, imaginativeness, general impression, and appraisal.

Moore (1990) designed the *Ability to Compose Music Exercise* in order to rate the ability of high school instrumentalists to complete a begun melody and to compose a complete melody based on contrasting words and pictures. The five criteria for rating ranged from *no expression* to *great deal of expression*. Each point along the scale was explicitly defined.

In a 1991 study, Kratus rated 60 songs of 7-, 9-, and 11-year-old subjects for the purpose of determining the 10 most successful and 10 least successful songs. The first item of the two item scale rated *craftsmanship* and contained a seven-point Likert scale with the anchors specifically defined. In a separate study (1994), Kratus analyzed the compositions of 40 third grade subjects using both rating scales and content analyses. The first rating scale was for *tonal cohesiveness*, and the second for *metric cohesiveness*. The ratings of these items were based on a seven-point Likert scale which offered the definition of the criteria to be rated as well as description of the anchors for the scale.

Bangs (1992) adapted an open-ended rating form from an art/poetry study by Amabile (1983) for the purpose of assessing 37 third grade subjects' musical compositions. The *Dimensions of Judgment* assessment tool contained 19 five-point Likert scale items with no definition of the criteria for each item.

The purpose of Smith's study (1993) was to evaluate the compositions of 18 piano students who ranged in age from 6- to 12-year-olds. The compositions were examined for differences in use of musical materials, structural properties, and originality and expressiveness. Items contained descriptions for each point of the scale as well as a checklist of criteria to help the rating process.

Analysis of the Studies

As we have stated earlier, the analysis of items from the eight scales revealed two approaches to style and two approaches to content. Although each scale as a totality seemed to conform to one construction approach or the other, all scales had mixtures of approaches.

The two style types might best be termed as *explicit* and *implicit*. Explicit items used more lengthy descriptions of what was to be rated and some actually offered criteria to consider. Implicit items carried little descriptive content, remaining purposefully vague in order for the evaluator to decide on meaning and criteria. Figure 1 displays examples for both principles. Examples 1 and 2 are clearly implicit in design, offering little explanation for words such as *pleasing* or *creative*. Examples 3-5 offer quite a different approach to design. In Example 3, anchors for a "7" and "1" rating are used. In Example 4 this is carried further with each item in the ranking scale described. Finally the last item explicitly suggests criteria that might be observed in order to define *syntactical logic*. Scales by Webster (1977, 1989), Kratus (1991, 1994), Moore (1990), and Smith (1993) all used explicit design. Bangs (1992) and Hassler and Feil (1986) used implicit design.

The two types of content centered on whether the judge was to consider specific music characteristics or more global issues. For instance, examples 1, 3 and 4 in Figure 1 ask the judge to consider a specific music characteristic and rate it accordingly. The presence of music characteristics included items for rhythm, harmony, texture, timbre, and expression. Not all studies dealt with all elements, but each had some attention to this kind of content. One study (Hassler and Feil, 1986) also included a chance for the evaluation of imaginative use of these music elements.

Examples 2 and 5 clearly address much broader matters considered to be more global. Global issues included items that best fit into the following four categories:

- originality/novelty/uniqueness
- general appeal/liking
- detail/elaboration
- syntax

Figure 1

Examples of Scale Construction and Content

Example 1 (Bangs, 1992)

- The degree to which there is a pleasing use of sound in the design.

5 4 3 2 1

Example 2 (Bangs, 1992)

- Using your own subjective definition of creativity, the degree to which the composition is creative.

5 4 3 2 1

Example 3 (Kratz, 1991)

- Tonal Cohesiveness - the degree to which the pitches in a composition are constructed around a tonal center or centers. (7 = very strong tonal cohesiveness, 1 = no tonal cohesiveness).

7 6 5 4 3 2 1

Example 4 (Smith, 1993)

- Finality of last cadence (Circle the number).
 - 5 - Strong sense of permanent conclusion.
 - 4 - Sense of conclusion, but a little unpredictable, given the implied harmonic context
 - 3 - somewhat inconclusive
 - 2 - weak sense of conclusion, difficult to predict
 - 1 - No sense of conclusion. seemingly random ending

Example 5 (Webster, 1989)

- Listen for the syntactical logic of the performance. Consider the following:
 1. Sensitivity to the creation of distinct parts
 2. Feeling of logical movement from one large event or set of events to another
 3. Return to a motive heard before
 4. Elaboration through sequence and/or repetition or a rhythmic idea or melodic contour
 5. Musical phrasing, with spots of relative repose
 6. Complimentary rhythmic or melodic motion
 7. Sensitivity to dynamics in relation to the whole
 8. Awareness of instrument tone quality and this awareness used to shape the piece musically
 9. Feeling of musical climax
 10. Sense of overall form
 11. Other musical aspects that contributed to syntactical logic

Rate the child's performance in terms of syntax. For ratings of "4" or higher, briefly note the qualities that serve as the basis for your rating:

5 4 3 2 1

Bangs (1992) also included global items for craftsmanship, creativity, and aesthetic value. Interestingly, these were the overriding constructs that we intended to use for concurrent validity. By embedding these items in the context of the rating scale forms, we could also see how they might relate to our judges' overall evaluations of the compositions after the rating scales had been completed.

Procedure

Research Questions

Given this analysis, we reasoned that in order to study these style and content issues we needed to create forms that paralleled one another in content, but were either explicit or implicit in style. We also reasoned that each form should be used by the *same* set of judges as they evaluated the *same* set of children's compositions, but that each set of forms would be used separately in time from one another. This would allow us to investigate interjudge reliability in several ways. We also decided to investigate how portions of each form might work to predict the overriding constructs of craftsmanship, originality/creativity, and aesthetic value by asking the same set of judges to make a final determination about the best and worst compositions in terms of these overall constructs as the last step in their work.

We set out to answer three major questions consistent with our purpose:

(1) Using one set of ten children's compositions and four expert judges, what are the interjudge reliability coefficients of two sets of rating scales, one constructed with explicit design and one with implicit design?

(2) Assuming adequate reliability as demonstrated in (1), how do the subscores of explicit/global (EG), explicit/specific (ES), implicit/global (IG), implicit/specific (IS) relate to and predict the global ratings of craftsmanship, originality/creativity and aesthetic value?

(3) How do judges' nominations of the *best* and *worst* compositions in terms of craftsmanship, originality/creativity, and aesthetic value relate to: (a) rating scales meant to judge the same thing, and (b) the EG, ES, IG, and IS subscores?

Method

Two separate rating forms were con-

structed based on the design of the items from the rating scales used in the eight studies described above. One form (Form I) included only those items that were implicit in design (see Appendix A). The second form (Form E) contained items that were exclusively explicit in design (see Appendix B). Each form contained parallel items separated into two large content categories: specific musical characteristics (S) and global considerations (G). Both forms had some items that fit into each of these categories. In a small number of instances where there were not parallel items for each form, we constructed items to fit the form. The organization of the two forms yielded subscores of IS (implicit form, specific musical characteristics), IG (implicit form, global considerations), ES (explicit form, specific musical characteristic) and EG (explicit form, global considerations).

Ten fifth and sixth grade children's compositions were selected from a pool of 24 compositions generated from two previous research studies (Hickey, 1993). In both studies, the subjects were given 30 minutes of free time to work toward creating an original composition using a Yamaha SY-55 synthesizer. No parameters were given except that they be prepared to record their composition after the 30 minute planning/practice period had expired. The final compositions were captured in MIDI file format using computer programs that allowed the recording of multiple tracks. The students were allowed to re-record their compositions as often as needed until they were satisfied with their final product.

Four independent, expert judges, all of whom were music educators with extensive teaching experience, were given an audio tape of the ten compositions and were asked to rate each of the compositions two times — once using rating Form I, and once using rating Form E. Two judges used rating Form I first, while the other two judges used rating Form E first. The judges were directed to listen to all of the compositions once before doing any kind of rating. All judges were asked to seal their first rating forms in an envelope immediately following the rating of the compositions, and to take a break of at least one

Table 1
Interjudge Reliability Correlations for Implicit (I) and Explicit (E) Forms
Using Specific Musical Content Scores

Judge	1		2		3	
	I	E	I	E	I	E
2	.67	.84				
3	.94	.83	.78	.81		
4	.74	.71	.80	.82	.84	.66

Table 2
Interjudge Reliability Correlations for Implicit (I) and Explicit (E) Forms
Using Global Content Scores

Judge	1		2		3	
	I	E	I	E	I	E
2	.81	.60				
3	.85	.73	.68	.49		
4	.87	.69	.92	.71	.80	.49

Table 3
Average Interjudge Reliability Correlations for Implicit Global (IG), Implicit
Specific (IS), Explicit Global (EG), and Explicit Specific (ES) Subscores

	G	S
I	.82	.80
E	.62	.78

hour before rating the compositions a second time using the second form.

To help answer research questions 2 and 3, the judges were also asked to fill out a brief third form (see Appendix C) asking them to choose what they felt were the *best* and *worst* two compositions in the categories of craftsmanship, originality/creativity, and aesthetic

value. A point system was devised that allowed the subsequent ordering of each of the ten compositions in each of the categories. Judges were also asked to comment qualitatively about their reactions to the separate forms.

The ratings for every item from each judge and for each composition were recorded for statistical

Table 4
Simple Regression of Subscores onto the Overriding Constructs of
Craftsmanship, Originality/Creativity, and Aesthetic Value.

Subscores	R	Adjusted R ²	t-value
Craftsmanship			
IS	.89	.76	5.37*
IG	.92	.83	6.81*
ES	.97	.93	10.65*
EG	.93	.84	7.01*
Originality/Creativity			
IS	.85	.68	4.48*
IG	.98	.95	13.59*
ES	.94	.87	7.89*
EG	.97	.93	11.20*
Aesthetic Value			
IS	.95	.90	9.02*
IG	.98	.95	12.63*
ES	.96	.91	9.59*
EG	.95	.90	8.65*
* $p < .01$.			

analysis. Correlations, t-tests, and simple regression were used in the analyses of the data.

Results

Research Question 1

Tables 1 and 2 show the average interjudge reliability between the ratings of the four judges on the two separate forms (I, E), for composite scores of specific musical items (S), and composite scores of global items (G). These results show relatively high interjudge reliability for both forms and for both specific music items and global items. We note a tendency for the Implicit coefficients to be higher, especially for the global items. Using these results, grand averages

for the subscores were computed and are displayed in Table 3. These results show that interjudge reliability for both musical and global items is higher on the implicit (I) form. This difference was statistically significant for the global content items (paired t test, $t = 8.1$, $p < .05$, $df = 5$).

Research Question 2

Table 4 displays the contribution each composite subscore (IS, IG, ES, EG) made to the overriding constructs of craftsmanship, originality/creativity, and aesthetic value which the judges did last (Appendix C). All of the subscores contributed significantly in their own way to the global ratings. It is interesting to note those subscores that contrib-

Table 5
Spearman Rank (rho) Correlation Matrix of Overriding Constructs with
Rating Scale Scores and Subscores of EG, ES, IG, and IS.

Overriding Constructs	Rating Scale Scores			Subscores			
	Crafts	Creat.	Aesth.	EG	ES	IG	IS
Craftmanship	.96*	.71	.80	.88*	.87*	.81	.82
Originality/Creativity	.70	.92*	.82	.81	.77	.89*	.80
Overall Aesthetic	.66	.70	.82	.62	.64	.72	.82
* $p < .01$.							

uted the highest in each category. For instance ES clearly explains nearly all of the variance for craftsmanship, while IG explains nearly all of the variance for originality/creativity and aesthetic value. This makes conceptual sense since explicit, specific ratings would tend to relate to overall craftsmanship and implicit global ratings would be associated with originality/creativity and overall aesthetic value.

Research Question 3

As Table 5 shows, the overriding constructs of craftsmanship and originality/creativity which the judges provided at the end of their work correlated most significantly ($p < .01$) with similar rating scale scores (as used in Bangs, 1992) within the rating scales. This suggests concurrent validity for these items. The lack of significance for the aesthetic value construct is puzzling. The construct for originality/creativity correlated significantly with global characteristic scores from the implicit form (IG), while the construct of craftsmanship correlated most significantly with scores from the more defined, explicit forms (EG, ES).

Discussion and Implications

It might be reasoned by some researchers that explicitly designed rating scales have greater reliability because the judge is given a clear sense of the item's meaning. Results

of this study using four judges' ratings of ten children's compositions suggest otherwise. Overall findings seem to suggest that approaches to rating scales that use consensual assessment as outlined by Amabile (1983) and others is, in fact, a profitable avenue for music teachers and researchers interested in children's composition.

It was of great interest to us that the rating scale which contained items that were very open-ended and implicit in nature (Form I) was in fact extremely reliable, and in the case of the global music content items, significantly more reliable than subscores from the form which was explicit (Form E). One judge commented that "... the subjectivity of this form [Form I] made it easier and more comfortable to me [in rating the children's compositions]. I felt as though I had artistic license to make professional judgments as I saw fit. The word 'pleasing' allowed me to listen with an aesthetic ear, rather than counting the number of timbres or textures I heard as I felt obligated to do with the first form [Form E]" Other judges noted similar preferences for the implicit form as well. One implication for further research is that time spent designing complex rating scale items might not return the dividends expected.

Many researchers include ratings of specific music characteristics as part of their ap-

Overall findings seem to suggest that approaches to rating scales that use consensual assessment as outlined by Amabile (1983) and others is, in fact, a profitable avenue for music teachers and researchers interested in children's composition.

proach to the study of music quality. Although this might be useful for other reasons, such ratings seem not to necessarily contribute to a better understanding of the constructs of originality/creativity and aesthetic value. Regression data appeared to indicate that using global rating items seem to explain the variance in all of the construct ratings equally as well. The subscores of global ratings from the implicit form (IG) prove to be most predictive for the constructs of the originality/creativity qualities and aesthetic value of children's compositions. The explicit form subscores were most predictive for the construct of craftsmanship, however, making this type of design perhaps more appropriate for rating of this specific characteristic.

Finally, it seems evident from the data that nominated *best* and *worst* compositions using the three overriding constructs created rankings that compared highly with both the subscores and internal rating scales designed to measure the same qualities. This provides some evidence for the concurrent validity of ratings scales such as those profiled here, and is quite encouraging for future research. It is interesting to note the significant correlation of both subscores from the explicit form (ES, EG) with rankings for craftsmanship whereas ratings from the implicit form (IS, IG) correlated more highly with the rankings for originality/creativity and overall aesthetic value. Perhaps when designing a rating scale for music compositions, researchers need to ask: is a child's craftsmanship, originality/creativity, or overall aesthetic value being assessed? Based on the results from this study, the careful design of the rating scale items and content should depend on the answers to these questions.

Another implication from this study is that judges can arrive at clear judgments of musi-

cal quality without formal musical analyses using explicit rating scales. This may not always be desired, particularly if specific information about craftsmanship is desired; however, the freedom to rely on the wisdom of expert judges without cluttering their considerations with cumbersome language and complicated design should be of interest to all. Further study of this kind should be pursued in the evaluation of complex music behavior.

References

- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997-1013.
- Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.
- 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, Florida.
- Boardman, E. (ed.). (1989). *Dimensions of musical thinking*. Reston, VA: Music Educators National Conference.
- Christensen, C. B. (1992). Music composition, invented notation and reflection: Tools for music learning and assessment. (Doctoral dissertation, Rutgers University, New Jersey, 1992). *Dissertation Abstracts International*, 53(06), 1834A.
- Cohen, V. W. (1980). The emergence of musical gestures in kindergarten children. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1990). *Dissertation Abstracts International*, 41(15), 4637A.
- Davies, C. D. (1992). Listen to my song: A study of songs invented by children aged 5 to 7 years. *British Journal of Music Education*, 9, 19-48.
- DeLorenzo, L. C. (1989). A field study of sixth-grade students' creative music problem-solving processes. *Journal of Research in Music Education*, 37(3), 188-200.
- Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: John Wiley.
- Hassler, M. & Feil, A. (1986). A study of the relationship of composition/improvisation to selected personal variables. *Bulletin of the Council for Research in Music Education*, 87, 26-34.

- Hedden, S. K. (1992, April). *Qualitative assessments of children's compositions*. Paper presented at the Music Educators National Conference, New Orleans, LA.
- Hickey, M. (1993). *Connecting the process to the product—creative problem finding in music and its relationship to the final compositions of 10-year-old children*. Poster presented at the MENC North Central Division Conference, Minneapolis, MN.
- Hickey, M. (1995). *Qualitative and quantitative relationships between children's creative musical thinking processes and products*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- Hoffman, K., Hedden, S. K., & Mims, R. (1990). *Music compositional processes in children aged seven through nine years*. Paper presented at the meeting of the American Orff-Schulwerk Association, Denver, CO.
- Hoffman, M. K., Hedden, S. K., & Mims, R. (1991). *Compositional processes in young children*. Paper presented at the Symposium on Research in General Music, The University of Arizona, Tucson.
- Kratus, J. (1985). The use of melodic and rhythmic motives in the original songs of children aged 5 to 13. *Contributions to Music Education*, 12, 1-8.
- Kratus, J. (1989). A time analysis of the compositional processes used by children ages 7 to 11. *Journal of Research in Music Education*, 37(1), 5-20.
- Kratus, J. (1991). Characterization of the compositional strategies used by children to compose a melody. *Canadian Journal of Research in Music Education*, 33, 95-103.
- Kratus, J. (1994). Relationships among children's music audiation and their compositional processes and products. *Journal of Research in Music Education*, 24 (2), 115-130.
- Levi, R. G. (1992). A field investigation of the composing processes used by second-grade children creating original language and music pieces. (Doctoral dissertation, Case Western Reserve University, Cleveland, 1991). *Dissertation Abstracts International*, 52(08), 2853A.
- Loane, B. (1984). Thinking about children's compositions. *British Journal of Music Education*, 1(13), 205-231.
- Moore, B. R. (1990). The relationship between curriculum and learner: Music composition and learning style. *Journal of Research in Music Education*, 38(1), 24-38.
- Reimer, B. (1989). *A philosophy of music education* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Smith, J. P. (1993). *Qualities of prompted and unprompted compositions*. Unpublished manuscript, Northwestern University, Evanston, IL.
- Swanwick, K., & Tillman, J. (1986). The sequence of musical development: A study of children's compositions. *British Journal of Music Education*, 3(3), 305-339.
- Wang, C. (1985). *Measures of creativity in sound and music*. Unpublished manuscript.
- 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). Conceptual bases for creative thinking in music. In J. Peery, I. Peery, and T. Draper (Eds.), *Music and child development* (pp. 158-174). New York: Springer-Verlag.
- Webster, P. (1989). *Measure of Creative Thinking in Music (MCTM)*. Administrative Guidelines. Unpublished manuscript, Northwestern University, Evanston, IL.
- Webster, P. (1992). Research on creative thinking in music: The assessment literature. In R. Colwell (Ed.), *Handbook of research on music teaching and learning* (pp. 266-280). New York: Schirmer Books.
- Webster, P., & Richardson, R. (1992). Asking children to think about music. *Arts Education Policy Review*, 94(3), 7-11.
- Wiggins, J. H. (1992). *The nature of children's musical learning in the context of a music classroom*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

7

APPENDIX A**Judgment of Musical Compositions Form I**

Judge's Name _____

Date: _____

Subject ID Letter _____

General Directions: By now, you have listened to the 10 subjects' compositions at least once. We ask that you listen to each composition again using the following rating form. The form is designed in two parts. Each part uses a Likert-type rating scale with numbers from 5 to 1. Please interpret "5" as the highest rating and "1" as the lowest. Circle (or check where appropriate) the rating number that you feel is appropriate for each item. Use your own definitions of the items given and try to be consistent in your interpretation of these items from subject to subject.

Part 1 asks you to rate specific musical characteristics. Note that we ask you to consider the presence of certain musical characteristics and also the imaginative use of some of these characteristics. Part 2 asks that you consider certain global issues about the compositions as a whole.

Part 1 Specific Musical Characteristics**Musical Characteristics (presence)****Rhythm**

- The degree to which the composition shows a pleasing use of rhythm.

5 4 3 2 1

Texture

- The degree to which the composition shows a pleasing use of texture (use of more than one instrument or pitch at a time.)

5 4 3 2 1

Timbre

- The degree to which there is a pleasing use of sound in the design.

5 4 3 2 1

Harmony

- The degree to which there is a pleasing use of harmony in the composition.

5 4 3 2 1

Expression

- The degree to which the work conveys dynamics, tempo, or high/low contrasts.

5 4 3 2 1

Musical Characteristics (Imagination)

- Imaginative treatment of the following musical elements: (Please place a check mark under the number of your rating for each element).

	5	4	3	2	1
Rhythm	—	—	—	—	—
Melody	—	—	—	—	—
Range (sound space)	—	—	—	—	—
Harmony	—	—	—	—	—
Expression	—	—	—	—	—
Form	—	—	—	—	—

Part 2 Global Considerations (Form I continued)

- First impression.

5 4 3 2 1

- Imaginative varying and ornamenting.

5 4 3 2 1

- In general, the degree to which the composition has aesthetic value.

5 4 3 2 1

- The amount of detail in the composition.

5 4 3 2 1

- The degree to which the composition conveys a sense of originality.

5 4 3 2 1

- The degree to which the composition displays craftsmanship.

5 4 3 2 1

- The degree to which the composition exhibits some unifying feature (i.e. motif, rhythm, melody, etc.).

5 4 3 2 1

- The degree to which the composition itself shows novel musical ideas.

5 4 3 2 1

- The degree to which the composition shows novel use of the instruments.

5 4 3 2 1

- The degree to which the composition shows variety.

5 4 3 2 1

- The level of complexity of the composition.

5 4 3 2 1

- Using your own subjective definition of creativity, the degree to which the composition is creative.

5 4 3 2 1

- Using your own subjective reaction to the composition, the degree you liked it.

5 4 3 2 1

APPENDIX B**Judgment of Musical Compositions Form E**

Judge's Name _____

Date: _____

Subject ID Letter _____

General Directions: By now, you have listened to the 10 subjects' compositions at least once. We ask that you listen to each composition again using the following rating form. The form is designed in two parts. Each part uses a Likert-type rating scale with numbers from 5 to 1, 7 to 1 or 3 to 0. Please interpret "5", "7" or "3" as the highest rating and "1" or "0" as the lowest. Circle the rating number that you feel is appropriate for each item. Use your own definitions of the items given and try to be consistent in your interpretation of these items from subject to subject.

Part 1 asks you to rate specific musical characteristics. Note that we ask you to consider the presence of certain musical characteristics and also the imaginative use of some of these characteristics. Part 2 asks that you consider certain global issues about the compositions as a whole.

Part 1 Specific Musical Characteristics**Musical Characteristics (presence)****Rhythm**

- Metric Cohesiveness - the degree to which the durations in a composition are constructed of regularly occurring accented and unaccented beats. (7= very strong metric cohesiveness, 1 = no metric cohesiveness).

7 6 5 4 3 2 1

- Tempo stability (Circle the number).

5 - pulse is very steady throughout the song
 4 - steady pulse wavers only on an occasional note
 3 - steady pulse wavers on a section
 2 - steady pulse occurs only occasionally
 1 - no sense of steady pulse at any point in the song

Texture

- The degree to which texture is used as a compositional device. 5 = variety of textures,
 1 = only one kind of texture throughout.

5 4 3 2 1

Timbre

- The degree to which different timbres are used as a compositional device. 5 = variety of timbres, 1 = only one timbre throughout.

5 4 3 2 1

Harmony

- Tonal Cohesiveness - the degree to which the pitches in a composition are constructed around a tonal center or centers. (7 = very strong tonal cohesiveness, 1 = no tonal cohesiveness).

7 6 5 4 3 2 1

- Finality of last cadence (Circle the number).
 - 5 - Strong sense of permanent conclusion.
 - 4 - Sense of conclusion, but a little unpredictable, given the implied harmonic context
 - 3 - somewhat inconclusive
 - 2 - weak sense of conclusion, difficult to predict
 - 1 - No sense of conclusion seemingly random ending
- Harmony (Circle the number).
 - 5 - use of harmonization appropriate to the musical context
 - 4 - some use of harmonization, often appropriate to the musical context
 - 3 - use of harmonization not necessarily implied by the musical context
 - 2 - use of harmonization only at cadences
 - 1 - no use of harmonization

Expression

- Check if present

<input type="checkbox"/> musical phrasing	<input type="checkbox"/> sensitive use of accelerando or ritard
<input type="checkbox"/> sensitivity to dynamics	<input type="checkbox"/> feeling of climax
<input type="checkbox"/> overall sense of form	<input type="checkbox"/> pianistic sense of style
<input type="checkbox"/> other musically expressive aspects	

Using a rating scale of 5 to 1 with 5 as the highest and 1 as the lowest, rate the composition for overall expressivity. (Circle number).

5 4 3 2 1

Comments:

- Expressive manipulation of melody, rhythm, articulation, dynamics, and form. (Circle the number).

1 = *No expression*. The composition did not use the specific musical element in questions (i.e. no articulation or dynamics were indicated).

2 = *Little expression*. The musical element was developed or used only once or twice (i.e., use of the same rhythm pattern throughout the entire piece or only one instance of a dynamic marking).

3 = *Some degree of expression*. Occasional manipulation or use of an element (i.e., occasional use of contrasting dynamic or articulation markings).

4 = *Good degree of expression*. Consistent expressive manipulation or use of an element throughout the piece (i.e., motivic use of rhythm or melody).

5 = *Great deal of expression*. Sophisticated means of development and elaboration (*i.e., retrograde, inversion, extended sequences) as well as a high level of expressiveness that can involve more than one element (*i.e., rhythm and melody manipulated together in parallel).

Part 2 Global Considerations (Form E continued)

- Check if present:

☐ unusual or changing meters ☐ unusual changes of direction
☐ large or frequent dynamic changes ☐ marked rhythmic complexity
☐ changing tempos musical ☐ other imaginative aspects

Using a rating scale of 5 to 1 with 5 as the highest and 1 as the lowest, rate the composition for overall originality. (Circle number)

5 4 3 2 1

Comments:

- Rate the extent to which the rhythm, melody, harmony, tone color, and texture has been elaborated. (Circle your choice).

3- Extensive Elaboration 2- Some Elaboration 1- Little Elaboration 0- No Elaboration

- Craftsmanship: 7 = the song forms a cohesive whole and makes interesting use of melodic and rhythmic patterns; 4 = the song has a moderate level of cohesiveness, and has some interest; 1 = the song appears to have no structure, with seemingly random pitches and rhythmic durations.

7 6 5 4 3 2 1

- Originality. Rate in regard to the uniqueness of expression. (Circle your choice).

3- Marked Uniqueness 2- Somewhat Unique 1- Little Uniqueness 0- No Uniqueness

- Listen for the syntactical logic of the performance. Consider the following:

1. Sensitivity to the creation distinct parts
2. Feeling of logical movement from one large event or set of events to another
3. Return to a motive heard before
4. Elaboration through sequence and/or repetition or a rhythmic idea or melodic contour
5. Musical phrasing, with spots of relative repose
6. Complimentary rhythmic or melodic motion
7. Sensitivity to dynamics in relation to the whole
8. Awareness of instrument tone quality and this awareness used to shape the piece musically
9. Feeling of musical climax
10. Sense of overall form
11. Other musical aspects that contributed to syntactical logic

Rate the child's performance in terms of syntax. For ratings of "4" or higher, briefly note the qualities that serve as the basis for your rating:

5 4 3 2 1

- Listen for unusual musical aspects of the performance. Consider:

1. Changing and/or unusual meters
2. Large and/or frequent dynamic contrasts
3. Changing tempi
4. Unusual use of direction change
5. Unusually large and/or small intervals
6. Marked rhythmic complexity
7. Unusual use sounds
8. Unusual musical combination and/or interchange between instruments
9. Other musical aspects that seem unusual or particularly imaginative

Rate the child's performance in terms of originality. For ratings of "4" or higher, briefly note the qualities that serve as the basis for your rating:

5 4 3 2 1

APPENDIX C

Overall Reactions

Judge: _____

Simply list the subject ID letters that you believe to be the 2 “best” and 2 “worst” in each of the following categories. A subject may be listed in more than one category.

1. Craftsmanship (technical skill).

Best 2

Worst 2

2. Originality/Creativity (imaginativeness).

Best 2

Worst 2

3. Overall aesthetic value (feelingful musical experience).

Best 2

Worst 2
