

THE DEVELOPMENT OF A RATING SCALE FOR JUDGING CONSTRUCTIVE FEEDBACK FOR STUDENT COMPOSITIONS

Maud Hickey
Northwestern University

Sam Reese
University of Illinois at Urbana-Champaign.

The purpose of this project was to develop a valid and reliable rating form for assessing the quality of written feedback university student mentors provide to school-aged students regarding their musical compositions. A second purpose was to develop a measurement tool for use in research related to effective composition mentoring. University students mentored school-aged students via the Internet. An initial form was developed and subjected to several analyses in order to create a reliable, valid, and efficient rating form. After collecting several items of good quality written feedback, 14 items were chosen and subjected to item analyses. Eight items were selected based on the item analyses and the final form was tested for interjudge reliability. The final interjudge reliability, based on four judges' ratings of 34 sample critiques was .74. This was lower than the reliability of .89 from the first interjudge reliability test which was based on 3 judges' ratings of 16 sample critiques.

Distance education, telementoring, and telecommunication technologies offer intriguing possibilities for connecting classroom teachers and preservice teachers with students via the Internet. No longer is it necessary for teachers and students to interact only in the same place at the same time. This sets up new possibilities for preservice teachers to gain teaching experience by working at a distance with real students in classrooms. However, this aspect of computer technology as a teaching tool also brings new problems for teacher educators in the assessment of preservice teaching skills. Though we have long been equipped for student teacher feedback in live classroom situations (e.g., Doerksen, 1990; Doyle, 1983), we are now confronted with a new mode of teaching via the Internet, which will require different forms of evaluation and assessment.

Two examples of Internet-based educational outreach exist at each of the authors' institutions. In each situation, preservice music education students (mentors) are connected via the Internet with public school students who are involved with technology-based music composition. The preservice mentors listen to original student compositions which are sent to them by students using web or conferencing software, then provide written con-

structive feedback to the students using an electronic form of communication similar to E-mail. This process of providing written feedback to students about their compositions is a basic part of the process of teaching composition, but differs from the typical form of spontaneous verbal critiques which music teachers provide to students when teaching in a live classroom situation.

In music education, preservice teachers are most likely trained to aurally detect musical errors in performing groups or classrooms and provide immediate verbal or musical feedback to students. They are rarely asked, however, to listen to original student compositions, analyze strengths and weaknesses, then provide constructive feedback to the student. This is due largely to the fact that in the United States, music composition as a school activity is quite new and is receiving fresh emphasis due in part to the inclusion of composing as a basic component of the National Standards for Music Education (*The School Music Program*, 1994). If our profession is to make real progress in helping students reach these new standards related to composing, teacher training institutions must begin to address the training and evaluation of preservice music teachers in the teaching of music composition as well as in using nontraditional forms of teaching and communication. Preservice teachers need instruction and practice in giving feedback on composition, and teacher educators need tools to assess the quality of their feedback.

While telecommunication technology and other forms of distance learning have been around for more than a decade in teacher education, it seems the majority of research has been focused on connecting teachers to each other, preservice teachers to each other, or teachers to preservice teachers and university instructors by way of E-mail or listserv communication (Burlbaw, 1993; Merseth, 1991; Pilburn & Middleton, 1998; Schlagel, Trathen, & Blanton, 1996; Thomas, Clift, & Sugimota, 1996; Willis & Mehlinger, 1996).

Pilburn and Middleton (1998) examined patterns of interaction between preservice teachers in collaborative reflective activities via a listserv and compared this to journal dialogue. Results suggest that listserv dialog has characteristics very different from typical classroom language. Electronic communication is very different from classroom discourse and deserves reexamination and careful attention.

Research on telementoring (on-line or Internet mentoring) is beginning to emerge, but is relatively scant (Willis & Mehlinger, 1996). The "Electronic Emissary," a large Internet-based telementoring and research effort, serves K-12 students, teachers, and experts from around the world (Harris, 1996). It probably offers the most extensive range of research and teaching resources related to telementoring to date. The project connects expert mentors with teachers and students and is driven by student or teacher curricular needs. The "HP Telemotor Program" allows professionals worldwide to help students, through teacher-supervised projects, in the critical areas of mathematics, science, professional communication skills, and career and education planning (Durkin & Neils, 1996). The "Learning Through Col-

laborative Visualization Project" (O'Neill, Wagner, & Gomez, 1996) is dedicated to connecting students and teachers with experts in various areas of scientific research. These successful efforts connect professional mentors to students in schools—not preservice teachers and students.

In music, most Internet or distance learning projects are between teachers or preservice teachers rather than between student and mentor (Walls, 1997). The "Vermont MIDI Distance Learning Network" (Cosenza & MacLeod, 1998), NETCOMM (Reese, 1999a), and MICNET (Hickey, 1999) are three examples of successful music telementoring projects which connect teachers, students, and composers in music composition. The projects are just beginning to produce data related to effective mentoring by preservice teachers (Reese, 1999b; Reese & Hickey, 1999).

In the present study we sought to develop a valid and reliable rating form for assessing the quality of written feedback that university student mentors provide to school-age students regarding their musical compositions. A second purpose was to develop a measurement tool for use in research related to effective composition mentoring. We developed an initial rating form and subjected it to several analyses in order to create a reliable, valid, and efficient rating form.

Method

The first phase of this study involved writing initial items for a rating form. To begin, the authors selected what they thought were excellent models of good feedback. We selected models which were written by the professional composers and exemplary preservice students involved in our respective pilot programs. Then the authors, along with experienced music teachers and composers, studied the selected pool of feedback critiques with the question: What makes these samples excellent models of constructive feedback? This was done in discussion format with the authors taking note of the answers to this question. The discussion results were then collected, organized, and arranged into descriptive statements. The process of organization involved deleting similar comments, combining those that logically fit together, and finally creating descriptive statements from them. The statements were subsequently revised into the 7-point Likert-type response items which became the initial pool for the rating scale.

The initial pool of 24 items was then examined by the authors of this study, a group of experienced music teachers, and composers for preliminary assessment of clarity, wording, etc. Items were either rejected, rewritten, or revised until we agreed that they were clear and unambiguous. A rating form was then created which included 14 items to be analyzed for the purposes of this study. In order to test the items in the rating form, samples of written feedback about original student compositions were selected to be judged.

Each author individually selected 12 written critique samples (24 total) from his or her respective Internet project classified in the following manner: 4 samples considered to be *excellent*, 4 of *medium* quality, and 4 of

poor quality. The authors then exchanged the randomly ordered samples and rated the overall quality of each other's selections as *excellent*, *medium*, or *poor*. The samples whose quality was agreed upon by both authors were chosen to narrow the pool. After discussion about each sample, we selected a total of 16 that represented a variety of age levels of composers being critiqued as well as quality of critique. The breakdown of the final critique samples to be rated by judges were 5 *excellent*, 5 *medium*, and 6 *poor*.

Three judges were chosen who were familiar with the two Internet projects either as teacher-participants or observers. Each judge was asked to rate each of the critique samples using the initial 14-item rating form. Judges' instructions were:

After listening to the musical file and reading the corresponding feedback, please rate the extent to which the quality of the feedback meets the criteria in each item. When finished, please provide comments regarding this scale in the space at the end of this form.

The judges scored each critique on a 7-item Likert-type scale in which 1 was described as *not evident* and 7 was described as *clearly evident*. The judges listened to the composition of the composer and knew the grade level of the composer for whom the critique was written. The ratings for all 16 samples by each of the three judges then were submitted to reliability and item analyses.

The rating form was subjected to analyses for interjudge reliability, concurrent validity, and item analysis. Two item analyses were performed to delete items and make the form more reliable. (The Final Rating Form is shown in Figure 1). Interjudge reliability and item analyses were performed in the final step to confirm the reliability and strength of all items in the final form. The results of these analyses are described next.

Results

The interjudge reliability between the 3 judges for all of the items in the initial rating form were calculated using Hoyt's analysis of variance procedure (Guilford, 1973). The interjudge reliability coefficient was .89.

To examine the concurrent validity of each item in the initial rating form, the judges' ratings of each sample were compared to the ratings of *excellent*, *average*, or *poor* initially assigned by the authors. The correlation coefficients of items with their quality ratings ranged from .07 to .80 (see Table 1). These were considered when selecting items for the final rating form.

An item-remainder coefficient (Spector, 1992) was calculated for each item in order to select the strongest items for a final rating form. This analysis was repeated three times until the final item pool and alpha coefficient were satisfactory. The interjudge reliability and concurrent validity were also taken into account. Table 2 shows the first step item analysis and corresponding alpha. Any item which had a low item-remainder coefficient (the correlation coefficient between that item with all other items) or a high coefficient alpha if removed (higher than the total coefficient alpha) were

removed. In this first step, items 1, 3, 6, 9, and 14 were removed because of their weak contribution to the overall form as illustrated in the item analysis (Table 2).

To the rater: After listening to the musical file and reading the corresponding feedback, please rate the extent to which the quality of the feedback meets the criteria in each item. When finished, please provide comments regarding this scale in the space at the end of this form.

1 = Not evident 7 = Clearly evident

Tone and Format of Feedback:

1. Positive feedback is specific to the composition or composition process, rather than general ("empty").

• 1 • 2 • 3 • 4 • 5 • 6 • 7

2. The critique of any weak areas in the composition is specific.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

Musical/Technical Analysis in Feedback:

3. Feedback includes clear analysis/description of the important musical elements of the composition

• 1 • 2 • 3 • 4 • 5 • 6 • 7

4. Feedback provides musical (and/or technical) terms which are appropriate for the age level of the composer.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

Focus of Feedback

5. Message contains specific suggestions for change.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

6. The suggestions for change are musically appropriate for this piece.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

7. Suggestions for change are appropriate for the age level of the composer.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

8. The writer uses effective devices to communicate imaginatively about suggestions or the piece as a whole, e.g. humor, metaphors, analogies, expressive language, etc.

• 1 • 2 • 3 • 4 • 5 • 6 • 7

Figure 1. Final Rating Form

Table 1

Item Correlation with Quality Rating

Items	<i>r</i>
Tone and Format of Feedback	
1. The overall tone is appropriately positive and constructive.	.38
2. Positive feedback is specific to the composition or composition process, rather than general ("empty").	.66**
3. The writer includes humor or "lightness" in the feedback which is effective and appropriate.	.07
4. The critique of any weak areas in the composition is specific.	.80**
5. The critique of any weak areas in the composition is honest.	.59*
6. The writer asks appropriate questions of the composer (e.g. related to intention, purpose or plans).	.14
Musical/Technical Analysis in Feedback	
7. Feedback includes clear analysis/description of the important musical elements of the composition	.63**
8. Feedback provides musical (and/or technical) terms which appropriate for the age level of the composer.	.63**
9. The writer uses effective metaphors or analogies to support given suggestions, to explain musical ideas, or to describe or suggest the expressive character of the piece.	.61*
Focus of Feedback:	
10. Feedback contains specific suggestions for change.	.72*
11. The suggestions for change are musically appropriate for this piece.	.66*
12. Suggestions for change are appropriate for the age level of the composer.	.66**
13. The writer uses effective devices to communicate imaginatively about suggestions or the piece as a whole, e.g. humor, metaphors, analogies, expressive language, etc.	.70**
14. The writer effectively refers to related styles of music, composers or pieces to support suggestions or musical ideas.	.32

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

Table 3 shows the results of the second step item analysis on only those items remaining after the first step items were removed. A cursory glance at the results of this analysis would point to the removal of item 7 because of the low inter-item coefficient and high alpha if removed. An alternative approach at this point is to also check all of the relatively weak items against

the external criteria (overall quality rating) and the effect of their removal on interjudge reliability (Spector, 1992). The item with the lowest correlation (among those remaining) with the quality rating was item 5 (see Table 1). Item 5 was removed and the results of the item analysis were very positive for the remaining items as shown in Table 4. In addition, the removal of

Table 2

Item Analysis – Step 1

Item	Item-Remainder Coefficient	Alpha if Item Removed
1	.59	.88
2	.75	.87
3	.34	.89
4	.81	.87
5	.63	.88
6	.14	.91
7	.62	.88
8	.80	.87
9	.50	.87
10	.74	.87
11	.69	.87
12	.77	.87
13	.80	.87
14	.42	.89

Note: Interjudge reliability = .89

Table 3

Item Analysis – Step 2

Item	Item-Remainder Coefficient	Alpha if Item Removed
2	.71	.93
4	.89	.92
5	.75	.93
7	.61	.94
8	.65	.93
10	.87	.92
11	.84	.92
12	.85	.92
13	.66	.93

Note: Interjudge reliability = .93

Table 4

Item Analysis – Step 3

Item	Item-Remainder Coefficient	Alpha if Item Removed
2	.75	.92
4	.87	.91
7	.63	.93
8	.67	.92
10	.85	.91
11	.80	.91
12	.82	.91
13	.67	.92

Note: Interjudge reliability = .93

item 5 did not affect the interjudge reliability as much as the removal of item 7 would have. Interjudge reliability with items 2, 4, 7, 8, 10, 11, 12, 13 was .89. Interjudge reliability using items 2, 4, 5, 8, 10, 11, 12, 13 was .87. A test of internal consistency (Cronbach, 1990) for items 5 and 7 was performed to compare their consistency. The coefficient alpha for item 5 was .62 and for item 7 was .74. Therefore it was decided to remove item 5 rather than item 7 from the final form.

For a second test of reliability, the final, eight-item form was given to four music educators to rate 34 additional samples of mentor written feedback. The mentors were 17 college music education majors who rated one music composition by a middle school student at the beginning, and then again at the end of a semester teaching period. The judges, who were experienced and familiar with the Internet composition projects as either teachers or mentors, listened to the student composition and then judged the 34 samples, which were presented in a random order. The interjudge reliability, using an intraclass correlation technique, was .74.

Discussion

The purpose of this project was to create and refine a scale for rating the quality of feedback on student compositions written by preservice teachers. After collecting several items to describe good quality written feedback, we chose an agreed-upon 14 items which were subjected to item analyses. After eight items were selected based on the item analyses, the final form was tested for interjudge reliability. The final interjudge reliability, based on four judges' ratings of 34 sample critiques was .74. This was lower than the reliability of .89 from the first interjudge reliability test which was based on three judges' ratings of 16 sample critiques.

The interjudge reliability of the rating form may be improved in two ways: (a) by giving judges more explicit and consistent training, and (b) by defining more clearly the anchors in the rating scale. Though the judges were very familiar with the background of the projects, and hence had apparent understanding about the purposes of the critiques, further discussion with judges after they had rated the critiques a final time revealed that they misunderstood some of the scale items. In further discussion, it seemed that this misunderstanding was related to the item anchor terms of *not evident* or *clearly evident*. These anchors could be better clarified by providing clearer and more unique rubrics for each item.

As we gradually learn to use the new communication tools of the next century, we can leverage their power to facilitate teacher-student interaction and to promote increased teaching and learning of music composition in school music programs. To do this, we need to discover the best characteristics of asynchronous teacher-student interaction in order to encourage the development of increased student skill in music composition. A fundamental type of interaction we need to understand is how the teacher should provide constructive feedback to the students about how to improve and extend their original compositions. The development of the rating scale in this study was an initial effort to clarify the characteristics of good constructive feedback regarding original student compositions and to create a reliable tool for assessing preservice teachers' growth in this skill. It is only one of a variety of new instruments that will be needed to help teachers learn the process of teaching composition and to encourage teaching in distributed partnerships for learning based upon distant, asynchronous, collaborative networks.

References

- Burlbaw, L. M. (1993). Supporting teacher education with computer-based telecommunication systems. *Journal of Educational Technology Systems*, 22, 99-112.
- Cosenza, G., & MacLeod, S. (1998). Vermont MIDI distance learning network: a model for technology in classroom music. In S. D. Lipscomb (Ed.), *Proceedings of the Fifth International Technological Directions in Music Learning Conference* (pp. 137-138). San Antonio, TX: IMR Press. Website: <http://www.webproject.org/midi>
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Doerksen, D. P. (1990). *Guide to evaluating teachers of music performance groups*. Reston, VA: MENC.
- Doyle, K. O. (1983). *Evaluating teaching*. Lexington, MA: Heath.
- Durkin, S., & Neils, D. (1996). *The HP E-mail Mentor Program* [On-line]. Available: <http://mentor.external.hp.cm/>
- Guilford, J. P. (1973). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- Harris, J. (1996). *The electronic emissary* [On-line]. Available: <http://www.tapr.org/emissary/>
- Hickey, M. (1999). *MICNET* [On-line]. Available: <http://collaboratory.acns.nwu.edu/micnet/>

- Merseth, K. K. (1991). Supporting beginning teachers with computer networks. *Journal of Teacher Education*, 42, 140-147.
- O'Neill, D., Wagner, R., & Gomez, L. (1996) On-line mentors: experimenting in science class. *Educational Leadership*, 54, 3, 39-42.
- Pilburn, M. D., & Middleton, J. A. (1998). Patterns of faculty and student conversation in listserv and traditional journals in a program for preservice mathematics and science teachers. *Journal of Research on Computing in Education*, 31(1), 62-77.
- Reese, S. (1999a). *NETCOMM: Network for technology, composing and music mentoring* [On-line]. Available: <http://www-camil.music.uiuc.edu/netcomm/Default.html>
- Reese, S. (1999b, October). The integration of on-line composition mentoring into music teacher education. Paper presented at the Association for Technology in Music Instruction National Convention, Denver, CO.
- Reese, S., & Hickey, M. (1999). Internet-based music composition and music teacher education. *Journal of Music Teacher Education*, (9), 1, 25-32.
- Sax, G. (1974). *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth Publishing Company.
- Schlagel, B., Trathen, W., & Blanton, W. (1996). Structuring telecommunications to create instructional conversations about student teaching. *Journal of Teacher Education*, 47 (3), 175-183.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-082). Newbury Park, CA: Sage.
- The school music program: A new vision*. (1994). Reston, VA: MENC.
- Thomas, L., Clift, R. T., & Sugimota, T. (1996). Telecommunication, student teaching and methods instruction: An exploratory investigation. *Journal of Teacher Education*, 47 (3), 165-173.
- Walls, K. C. (1997). Music performance and learning: The impact of digital technology. *Psychomusicology*, 16, 68-76.
- Willis, J. W., & Mehlinger, H. D. (1996). Information technology and teacher education. In J. Sikula, T. J. Buttery, and E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 978-1029). NY: Macmillan.