The Contours of Inclusion:
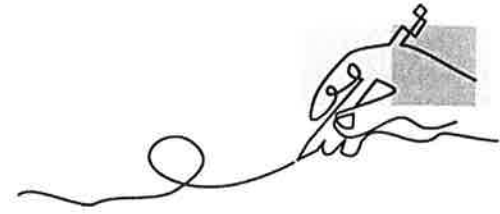## Frameworks and Tools for Evaluating Arts in Education

VSA arts

What You See Is What You Get: The Development of an Observational Strategy

Robert Horowitz, Ed.D.
Teachers College, Columbia University

During the 2005–2006 school year, ArtsConnection of New York City launched its DELLTA (Developing English Language Literacy Through the Arts) program. As one part of our DELLTA[1] program evaluation, our research team set about developing an observational strategy that would build upon our prior work while also providing rich and accurate data from DELLTA's artist residencies.[2]

We began with five criteria for instrument development. The instrument needed to be unobtrusive, practical, authentic, valid, and reliable.

- **Unobtrusive**—Our presence (as researchers) in the classroom should have as little impact as possible on the artists' and teachers' ability to teach and the children's ability to learn.
- **Practical**—The instrument should be relatively easy to use by experienced researchers in varied classroom contexts. Data should be gathered, coded, aggregated, and analyzed as efficiently as possible.
- **Authentic**—Observational data should reflect authentic artistic and academic experiences.
- **Valid**—The assessment criteria in the instrument should accurately reflect the intended and operational instructional content of the artists and teachers.
- **Reliable**—The instrument should yield accurate and consistent data when used by different researchers in different schools and classrooms.

While these criteria are certainly not mutually exclusive, our efforts to make the instrument practical and unobtrusive had the potential for reducing its reliability. Our data would be more accurate if multiple raters assessed individual children, for instance, but then we would not be able to use it in numerous diverse and authentic classroom settings. Therefore, we attempted to strike a reasonable balance between potentially competing criteria.

To develop the ArtsConnection DELLTA assessment, we needed to choose a set of observable indicators that reflected DELLTA instructional content and was consistent with our prior and current research on ArtsConnection's programs. And because ArtsConnection's DELLTA program was implemented in both elementary and middle schools, and included dance and theater artists, we wanted to develop four assessment instruments to reflect each possible instructional and grade level possibility. However, we also wished—as much as possible—to use the same variables in each of the instruments so that we could

analyze data across different classroom settings.

To start, we gathered a potential set of observable student indicators. We used three sources:

- Criteria within ArtsConnection-developed Observation Sheets[3];
- Our prior research on aspects of student development within cognitive, social, and personal domains; and
- Our ongoing qualitative research within ArtsConnection residencies.

ArtsConnection's Observation Sheets are part of their professional development process within inquiry-based arts partnerships. They help artists and teachers learn to focus on their students' observable behaviors in dance and theater. Behavioral indicators adapted from the Observation Sheets included *physical control, coordination and agility*, and *spatial awareness* (for dance), and *physical awareness, physical expression*, and *commitment* (for theater).

We selected additional behavioral indicators by comparing our qualitative data with our prior research on cognitive, social, and personal development. In the *Learning In and Through the Arts* study at Teachers College, we identified a model of cognitive skills, social competencies, and personal dispositions that operated within arts learning that were also operational within other academic subjects and life experiences, and might serve as the mechanism of transfer between the arts and other areas.[4] Subsequently, in our research and evaluation of ArtsConnection partnerships, we developed a series of rating scales to assess aspects of the cognitive-social-personal model.[5] Indicators selected from this model included *elaboration, motivation, ability to focus, perseverance/task persistence*, and *ownership of learning.*

We also chose several indicators that reflected DELLTA program goals or areas of student development that we had observed, such as *acquisition of English language skills* and *gives constructive feedback to other students.*

The student behavioral indicators were matched with a set of pedagogical indicators reflecting superior instruction or collaboration by artists and/or teachers. These were also based upon our prior and ongoing research, and included areas such as *provides opportunity for verbal expression, effective collaboration between artist and teacher*, and *teacher support and buy-in.*

Four observation instruments were set up as four spreadsheet worksheets (for dance and theater, and for elementary and middle school). We named the instrument the Classroom Assessment for Learning and Teaching (CALT), distributed it to a team of four field researchers, and began to use it within a diverse group of DELLTA residencies. (See Tables 1 and 2.)

Each Excel worksheet had three columns (or fields) to submit data. In the first data field, the researchers estimated the percentage of students who demonstrated a behavior that indicated achievement within a certain indicator. Estimates could range from 0 to 10, representing 0% to 100%. In our study, researchers estimated the maximum achievement during any one moment in a class, although the protocol could also be used at specified intervals or to assess different classroom tasks. If the artist or teacher did not provide a classroom opportunity for achievement within a particular indicator, then that spreadsheet cell was left blank.

Researchers also noted if they observed the presence of the pedagogical indicators. We decided to make these dichotomous variables—that is, the researchers simply checked off whether or not they observed the artists or teachers demonstrate these behaviors.

In the second data field, researchers input text that described the behaviors representing each indicator. The researchers were encouraged to be as detailed as possible, and to describe individual children's behaviors rather than submit generalities about the entire class's behavior. These slices of qualitative data were essential for our analysis. We purposely did not select specific behaviors in advance to represent successful achievement within our indicators. Instead, we hoped to develop a body of qualitative data that described each indicator. This would help us refine the instrumentation in the future, and help us understand how the field researchers were responding to the observation instrument and the classroom experience. Equally as important, we expected that the qualitative data would help us describe to program participants the aspects of the teaching and learning experiences that were most successful, and those aspects that were not as effective. The third data field was labeled "notes" and provided researchers with the opportunity to additionally annotate their statistical estimations and their qualitative data.

In spring 2007, four researchers used the new instruments in a varied group of ArtsConnection dance and theater residencies in several schools in Brooklyn, Queens, and Manhattan, New York City. The instruments were used for 24 elementary school classroom observations and 6 middle school observations. Field researchers took detailed notes during each observation of an artist residency within a participating classroom. Later, they filled in the spreadsheet/protocol for each class they observed and submitted it to "ArtsResearch Central" via e-mail.

Table 1: CLASSROOM ASSESSMENT FOR LEARNING AND TEACHING—THEATER

Student Indicators—Elementary School Theater	Achievement	Behavioral Indicators	Notes
1. Physical awareness			
2. Physical expression			
3. Vocal expression			
4. Commitment			
5. Observation			
7. Creative expressiveness in theater			
8. Imagination/problem solving			
9. Verbal or written expression of ideas or feelings			
10. Application of vocabulary from arts classes			
11. Acquisition of English language skills			
12. Elaboration			
13. Cooperative learning skills/collaboration			
14. Motivation			
15. Perseverance/task persistence			
16. Ability to focus			
17. Ownership of learning			
18. Self-confidence/risk taking			
19. Demonstrates good audience skills			
20. Gives constructive feedback to other students			
Teaching Indicators (Check)			
1. Provides opportunities for verbal expression			
2. Provides opportunities for expression in theater			
3. Provides opportunities for application of vocabulary			
4. Makes explicit connections between theater and English skills			
5. Makes other academic or cultural connections			
6. Provides opportunities for editing, elaboration, or improvisation			
7. Fosters broader understanding of theater			
8. Supports effective collaboration between artist and teacher			
9. Teacher support and buy-in			
10. Promotes teacher comfort and confidence with using theater			
11. Fosters sense of ownership			
School:		Researcher:	
Teacher:		Date of observation:	
Artist:		Number of students observed:	

Table 2: CLASSROOM ASSESSMENT FOR LEARNING AND TEACHING—DANCE

Student Indicators—Elementary School Dance	Achievement	Behavioral Indicators	Notes
1. Physical control			
2. Coordination and agility			
3. Spatial awareness			
4. Observation and recall			
5. Rhythm			
6. Movement qualities			
7. Improvisation			
8. Creative expressiveness in dance			
9. Verbal or written expression of ideas or feelings			
10. Application of vocabulary from arts classes			
11. Acquisition of English language skills			
12. Elaboration			
13. Cooperative learning skills/collaboration			
14. Motivation			
15. Perseverance/task persistence			
16. Ability to focus			
17. Ownership of learning			
18. Self-confidence/risk taking			
19. Demonstrates good audience skills			
Teaching Indicators (Check)			
1. Provides opportunities for verbal expression			
2. Provides opportunities for expression in dance			
3. Provides opportunities for application of vocabulary			
4. Makes explicit connections between dance activities and English skills			
5. Makes other academic or cultural connections			
6. Provides opportunities for editing, elaboration or improvisation			
7. Fosters broader understanding of dance			
8. Supports effective collaboration between artist and teacher			
9. Promotes teacher support and buy-in			
10. Facilitates teacher comfort and confidence with using dance			
11. Fosters sense of ownership			
School:		Researcher:	
Teacher:		Date of observation:	
Artist:		Number of students observed:	

Analysis Process

The spreadsheet and Internet-based data collection method expedited the analysis process. A master spreadsheet was developed by aggregating the spreadsheet submissions from each researcher. Overall mean scores and standard deviations, for all researcher estimates, were calculated for each indicator. We also obtained scores for *opportunity*. This score indicated the percentage of observable opportunities for each indicator, because the researchers left cells blank if the class they observed did not provide an opportunity for estimating achievement for an indicator. We did not use estimates for an indicator if an *opportunity* score was less than 35%.

We obtained overall statistics, as well as scores for individual artists, schools, and researchers. We compared scores from different researchers to obtain initial estimates of inter-rater reliability.

By using the sorting features of the spreadsheet application, we were able to extract all of the qualitative data coded according to each behavioral indicator. Then we could examine all descriptions of children's behaviors for each indicator, and analyze by looking for patterns and the most salient behaviors. With a little experimentation, the spreadsheet proved to be effective for qualitative analysis and served some of the elementary functions of qualitative analysis software.[6]

Results

Results in the following tables show the average estimates for each student indicator. The indicators are placed in rank order of highest to lowest scores, to demonstrate the most salient indicators.

For example, the strongest score in theater was in student *motivation*, with an average of 83.85% of observed students (across all classes and observations) demonstrating behaviors indicating *motivation* within ArtsConnection DELLTA residencies. The second highest score was in *cooperative learning/collaboration*, with an average of 78.57% of students demonstrating this skill. It is interesting to note, however, that the standard deviation is much lower for *motivation* than for *cooperative learning* (8.7% compared with 23.4%), indicating more agreement among raters on the *motivation* variable.

Other high-scoring indicators in theater included *ability to focus* (78.18%; SD = 14.0), *perseverance/task persistence* (77.78%; SD = 9.7), and *commitment* 73.85%; SD = 13.3).

Table 3: CLASSROOM ASSESSMENT FOR LEARNING AND TEACHING—THEATER

Student Indicators— Theater	Mean	SD
Motivation	83.85%	8.7
Cooperative Learning/Collaboration	78.57%	23.4
Ability to Focus	78.18%	14.0
Perseverance/Task Persistence	77.78%	9.7
Commitment	73.85%	13.3
Imagination/Problem Solving	73.33%	15.8
Ownership of Learning	71.11%	13.6
Observation	68.18%	14.0
Physical Awareness	66.92%	13.8
Vocal Expression	66.36%	21.6
Physical Expression	65.83%	16.8
Creative Expressiveness in Theater	62.50%	18.6
Self-Confidence/Risk Taking	60.00%	25.5
Acquisition of English Language Skills	56.00%	24.1
Elaboration	50.00%	19.1

In Table 4, the strongest score in dance residencies was also in the area of *motivation* (89.00%; SD = 8.8). Other high-scoring indicators in theater included *improvisation* (82.50%; SD = 10.4), *ability to focus* (82.00%; SD = 11.4), *ownership of learning* (81.11%; SD = 12.7),[7] and *observation and recall* (80.00%; SD = 10.0).

By combining CALT scores from both the dance and theater residencies, we were able to identify the highest scoring indicators across arts disciplines, schools, and artists. *Motivation* was the most significant overall indicator, followed by *ability to focus, perseverance/task persistence, cooperative learning/collaboration*, and *ownership of learning*. These results indicate that our observers found that these were areas where ArtsConnection was strongest in supporting student growth through their arts residencies.

Table 4: CLASSROOM ASSESSMENT FOR LEARNING AND TEACHING—DANCE

Student Indicators— Dance	Mean	SD
Motivation	89.00%	8.8
Improvisation	82.50%	10.4
Ability to Focus	82.00%	11.4
Ownership of Learning	81.11%	12.7
Observation and Recall	80.00%	10.0
Creative Expressiveness in Dance	80.00%	10.0
Perseverance/Task Persistence	80.00%	10.5
Self-Confidence/Risk Taking	80.00%	21.6
Spatial Awareness	77.00%	10.6
Rhythm	73.00%	14.9
Physical Control	71.00%	13.7
Verbal or Written Expression of Ideas or Feelings	67.78%	19.2
Cooperative Learning/Collaboration	67.50%	18.3
Coordination and Agility	66.00%	15.1
Movement Qualities	64.29%	23.7
Acquisition of English Language Skills	43.33%	15.1

Strongest Indicators across Art Forms, Schools, and Artists

1. Motivation
2. Ability to Focus
3. Perseverance/Task Persistence
4. Cooperative Learning Skills/Collaboration
5. Ownership of Learning

Results from teacher surveys supported these findings. A series of rating scales was administered to participating teachers, reflecting learning in three of these five areas.[8] Teachers responded very positively, indicating their perceptions that these were areas of student growth supported by the arts residencies.

Teachers reported that students were more motivated to succeed because of the arts residencies. Almost all teachers responded positively to each item, with a majority strongly agreeing that "children accomplished more than expected, because they were challenged."[9]

Table 5: TEACHER REPORT—MOTIVATION

Motivation	SA	A	N	D	SD
Children accomplished more than expected, because they were challenged.	54%	43%	4%	0%	0%
Otherwise difficult students tried harder in the arts classes.	64%	29%	7%	0%	0%

SA = strongly agree   A = agree   N = not sure   D = disagree   SD = strongly disagree

Teachers reported that students developed cooperative learning skills through the residencies. A large majority of teachers "strongly agreed" that "students working in groups demonstrated good coordination, allowing each other to speak and try each other's ideas" and that they "realized they could work together on group arts projects in spite of their differences" and "put aside differences to reach a common goal."

Table 6: TEACHER REPORT—COOPERATIVE LEARNING

Cooperative Learning	SA	A	N	D	SD
Children in group work understood that they were not out there all alone, and that everyone could contribute.	82%	14%	4%	0%	0%
Students working in groups demonstrated good coordination, allowing each other turns to speak and try out each other's ideas.	68%	32%	0%	0%	0%
The children realized they could work together on group arts projects in spite of their differences.	75%	25%	0%	0%	0%
In groups, students could put aside their differences to reach a common goal.	64%	29%	4%	4%	0%

SA = strongly agree   A = agree   N = not sure   D = disagree   SD = strongly disagree

Teachers reported that students were learning to take charge of the learning process. A large majority of teachers "strongly agreed" that students' "work belonged to them, not to the teacher or artist" and that their work "reflected their personal experiences."

Table 7: TEACHER REPORT—OWNERSHIP OF LEARNING

Ownership of Learning	SA	A	N	D	SD
Students' work in dance or theater reflected their personal experiences.	64%	29%	4%	4%	0%
Students felt that they decided what was in their own artwork (in dance or theater).	48%	41%	7%	4%	0%
Students felt that their work belonged to them, not to the teacher or artist.	71%	18%	7%	4%	0%

SA = strongly agree   A = agree   N = not sure   D = disagree   SD = strongly disagree

Final Thoughts

Our team found several benefits to our use of the Classroom Assessment for Learning and Teaching (CALT) observational strategy:

- It helped focus our observations and note-taking on specific aspects of students' behaviors related to program goals and evaluation objectives. The data collection process was streamlined, our researchers' energies were focused, and we were able to gather and analyze a reasonably large amount of data with sufficient efficiency.
- This, in turn, helped us talk with ArtsConnection more effectively about what we were learning, and the aspects of classroom implementation that were most likely to lead to success. ArtsConnection could compare our results with their own internal assessments and observational strategies, make judgments about their own effectiveness, and better communicate their accomplishments to their own constituencies and stakeholders.
- The CALT data provided additional corroboration for findings derived from descriptive observation, teacher rating scales, and teacher interviews. This year we are applying the CALT strategy throughout more residencies, to determine if student behaviors change over time. The CALT data are building blocks in an overall evaluation design that includes matched teacher assessments (via rating scales), student assessment data, and coded qualitative data.

We also found that the CALT observation strategy posed several challenges:

- It is generally easier to use the CALT format to assess concrete skills in the arts than for abstract concepts such as *ownership of learning*. Our use of similar strategies in other projects was sometimes easier when we assessed whether children were achieving tasks such as singing in tune or playing a musical instrument in rhythm. However, it was interesting to us that constructs such as *ownership of learning* still rose to the top of our ratings.
- We worked with a tight-knit group of researchers with years of experience working together on our definitions of cognitive, social, and personal development. Newer researchers would need more training to reliably use the observation system.
- One field researcher felt that the system was "reductionist" as it forced us to oversimplify complex constructs that are best left to rich, qualitative investigation. However, this seems to be a problem inherent to quantification, which seeks to represent verbal (and sometimes ambiguous) concepts with a scaled set of numbers. Both the limitations and benefits of quantification were evident to our team. But the paired qualitative data—input to the spreadsheet and then recoded—helped us interpret and understand the statistical data.

We hope that others can borrow some of the ideas presented here to develop observational strategies relevant to their own programs. An arts partnership evaluation can benefit from an assessment process that is strongly rooted in the program's unique experience, and not simply based upon external, desirable criteria. At the very least, this process helped our researchers to better observe the distinctive aspects of the ArtsConnection DELLTA program. It provoked more substantial conversations with the program's participants, thereby improving instruction and helping us learn more about how to assess student learning.

Acknowledgments

Founded in 1979, ArtsConnection is New York City's most comprehensive arts in education organization. ArtsConnection believes that the arts are essential to education, and intrinsic to the social, cognitive, and personal development of every child. To realize this vision, ArtsConnection's faculty of 173 teaching artists work with classroom teachers at all grade levels (pre-K through 12) in creative collaborations that yield powerful arts learning experiences for children, teachers, and families. ArtsConnection continually refines its strategies to meet changing educational needs, bringing depth and diversity to artist residencies, performances, family and after-school programs in music, dance, theater and the visual arts. To maximize the success of its programs and build capacity for the arts in education, ArtsConnection also provides extensive professional development for teaching artists, classroom teachers, and arts specialists; as well as conducts research and evaluation to assess programs and share best practices with the field. These efforts have made ArtsConnection a full-service educational partner with the New York City Department of Education, providing more than 14,000 instructional hours in over 120 partner schools, reaching 30,000 participants annually.

---

[1] ArtsConnection has received funding from the U.S. Department of Education for the Developing English Language Literacy Through the Arts (DELLTA) project. DELLTA is a three-year project in five elementary schools (2005–2008) and a four-year project in three middle schools (2006–2010) that serves students who are designated English-language learners and their classmates. The methodologies are grounded in the processes of professional development and ongoing collaboration between teachers and artists that facilitate inquiry-based partnerships focused on the questions, "What is the nature of teaching and learning in dance and theater?" and "In what ways do they influence language acquisition in English-language learners?"

[2] Researchers contributing to this study included Elizabeth Beaubrun, Amy Kleiman, and Dr. Dan Serig.

[3] ArtsConnection's Observation Sheets were developed as part of the Young Talent Program (YTP) with grants from the Jacob K. Javits Gifted and Talented Students Education program at USED in the early 1990s. Categories for the Observation Sheets were developed by artists in collaboration with classroom teachers. ArtsConnection has used them for 15 years in the YTP and has adapted them for broader use in the DELLTA project. ArtsConnection artists have reported that they facilitate productive conversations with teachers because they are grounded in shared observation of student behavior in the arts. For a description of the development of the Observation Sheets, see S. Baum, S. Owen, and B. Oreck, "Talent Beyond Words: Identification of Potential Talent in Dance and Music in Elementary Students," *Gifted Child Quarterly* 40 (1996), 93–101.

[4] E.B. Fiske, ed., *Champions of Change: The Impact of Arts on Learning*. Washington, D.C.: Arts Education Partnership and President's Committee on the Arts and the Humanities (1999).

J. Burton, R. Horowitz, and H. Abeles, "Learning In and Through the Arts: The Question of

Transfer," *Studies in Art Education* 41(3), 228–257.

[5] R. Horowitz and A. Kleiman, "The Relationship Between Arts Learning and Cognitive Skills, Social Competencies, and Personal Dispositions." Paper presented at the meeting of the American Education Research Association, New Orleans (2002).

R. Horowitz, "Connections: The Arts and Cognitive, Social, and Personal Development," in *Partnering Arts Education: A Working Model from ArtsConnection,* ed. B. Rich (32–48). New York: Dana Press (2005).

[6] If you try this on your own, we suggest adding a number before each indicator in the first column. Then you can run a sort on that field and keep the data in the appropriate order. Once the qualitative data is sorted, you can paste the field into a word processing program and "convert table to text" to put the data in a more manageable format, similar to a report from a qualitative analysis package.

[7] To us, *ownership of learning* denotes a sense of children taking charge of the learning process, with their work distinctly their own and not simply the products of a teacher-directed classroom experience.

[8] The rating scale development process is described in this publication from the Dana Foundation: R. Horowitz, "Connections: The Arts and Cognitive, Social, and Personal Development," in *Partnering Arts Education: A Working Model from ArtsConnection,* ed. B. Rich (32–48). New York: Dana Press (2005).

[9] Rows in these tables might not total 100% due to rounding.