

Evaluation in the Arts Is Sheer Madness

[RICHARD COLWELL](#)

UNIVERSITY OF ILLINOIS

ABSTRACT

Arts educators have two opinions on evaluation: they are continually evaluating or they believe the important outcomes of their teaching defy systematic assessment. Assessment depends upon a clear definition of the discipline. Arts educators focused primarily on performance (production) do assess individual and group objectives in terms of product. Assessment within the reform movement (including standards) and assessment in curricula such as DBAE require a broader approach including a differentiation between program evaluation and evaluation to improve student competence. These and other assessment issues are raised as a means of initiating professional dialogue in contemporary arts assessment and the demands being placed upon the arts.

INTRODUCTION

Madness is associated with genius or artists as well as funny farm inmates. Sometimes it applies to educational ventures. Other definitions include “great folly” and “enthusiasm or excitement.” Today’s discussions about evaluation in the arts encompass all of these definitions and more. Evaluation specialists are providing us with challenging results based on research in language arts and mathematics. Non-specialists, including artists and arts supporters have seized upon evaluation as an avenue to promote their idea and/or cause. The primary purpose of this article is to clear-up some of the confusion associated with evaluation in the arts and to advance arguments that evaluation can facilitate teaching and learning in the arts. One approaches madness when one wades through the literature in arts education that relates to evaluation issues. You can find anything. A major problem is that the language used is imprecise. I’ll refer to it as “fuzzy”, fuzzy thinking and fuzzy language. Fuzzy language is important in foreign policy as public language and private meaning are both used; similarly, vagueness has been an aid to promoting the importance of the arts. Policy gurus in the arts and in education find imprecision helpful in that the context and/or situation will affect how policy is implemented. Precision is neither wanted nor valued.

Evaluation is usually associated with reporting the success and/or failure of a process or a product. A clear concept and a precise definition of that process or product are necessary to interpret the results of the evaluation. Evaluation and fuzziness do co-exist but not when it comes to understanding. Goal free evaluation, evaluation without prior objectives, can be extremely valuable; the evaluation being touted and condemned in 2004, however, is associated with objectives, aims, goals, a purpose, standards, and more. Accordingly, I will first wrestle with definitions and the context of arts education to explain how and why we find confusion in the field. This done, I will then devote some space to the relationship of standards and evaluation, discuss the reform movement and the policy thrusts provided by the No Child Left Behind legislation, briefly glance at colleges and teacher education, and, finally, focus on classroom evaluation, which should be the primary evaluation concern of arts educators. Educators are prone to distinguish among testing, measurement, assessment (both authentic and its opposite),

evaluation, and accountability. These important technical definitions are not a concern here except to distinguish between accountability and assessment. I use evaluation and assessment interchangeably. Measurement is thought to imply the use of objective and precise evaluation tools -- although the precision may focus on trivia. In music, the *Seashore Measures of Musical Talents* was considered the gold standard for precision in assessing aptitude, although its primary purpose was actually to provide a holistic glimpse of individuals who should not be encouraged to study music seriously. Thus, at the outset, let me offer the caveat that the field is messy, has been for a long time, and arts professionals continue to have difficulty communicating on the issues among themselves. Evaluation along the lines advocated by education psychologists has not been our forte. Conveying to educators and the public what we mean by arts assessment is difficult. The public sees performance; those educators in systematic evaluation don't understand us. The lack of attention to the technicalities of evaluation is easily explained in John Dewey's terms. Neither arts educators nor their students (including parents) have a felt need. Today's uneasy felt need, if one exists, is to defend the profession, not to change or improve it. Why devote valuable curricular time in teacher education or in the classroom on a topic that will not be used beyond "common-sense" assessments? There is no public outcry that students are not proficient in dance, theatre, the visual arts, or music. There is also no evident concern for the quality of today's arts educators other than the continuing concern that better teachers and better teaching are always needed.

An example of near-madness can be found in the attempts to define arts education clearly evident in the courses allowed by the various states to meet an arts requirement. Languages other than English may be "arts", literature or certain areas of history may be "art". Waivers are common ---what about out of school arts experiences? Similarly confused is the definition of arts education and an arts educator. To the best of my knowledge, there are few individuals, teaching in the schools educated as an arts educator to serve as a clarifying model. The Consortium of National Arts Education Associations (1994) defines the following "standards" as representing the competencies of a high school graduate who has received an adequate education in the arts. Presumably all teachers would exceed these standards.

“They should be able to communicate at a basic level in the four arts disciplines—dance, music, theatre, and the visual arts. This includes knowledge and skills in the use of the basic vocabularies, materials, tools, techniques, and intellectual methods of each arts discipline. They should be able to communicate proficiently in at least one art form, including the ability to define and solve artistic problems with insight, reason, and technical proficiency. They should have an informed acquaintance with exemplary works of art from a variety of cultures and historical periods, and a basic understanding of historical development in the arts disciplines, across the arts as a whole, and within cultures. They should be able to relate various types of arts knowledge and skills within and across the arts disciplines. This includes mixing and matching competencies and understandings in art-making, history and culture, and analysis in any arts-related project” (Consortium, 1994, 18-19).

In the years since 1994, the arts disciplines of dance, music, theatre, and visual arts seem to have ignored this definition. Each has developed subject-matter competencies or “standards” that bear only marginal relationship to the arts standards. Few teachers practicing at any level would be judged proficient or advanced on the arts “standards.”

A recent article by Jessica Davis is typical of the messiness of the relationship of the discipline of evaluation and the discipline of this amorphous “arts education” (Davis, 2003, 28, 30). She rightly laments the position of arts advocates who see the role of evaluation as that of documenting how the arts benefit school achievement, including attendance and motivation, and then proceeds to slam standardized tests (in any subject) en route to promoting the importance of *process* in arts education for all students. I admire Jessica Davis for reminding fellow educators of the importance of the arts and for her “solution” that all students might learn to handle failure through arts experiences. Her argument is useful primarily for advocacy and at the broadest levels of curriculum consideration. Not much on evaluation, the idea that a teacher can be educated in “arts education” is limited to a small number of colleges, primarily Lesley and Harvard Universities in the Boston area, and runs counter to the importance in education of subject matter knowledge. Not only would most music teachers feel

incompetent if assigned to teach a course in visual arts or dance, but most vocal musicians would even be hesitant to assume responsibility for an instrumental music program, whether strings or the band. I have in my test files a draft instrument developed by ETS for Oberlin College,-- undated, but probably from the 1960s, -- that expected music students to have a breadth of knowledge in visual art and music, a test that was quickly dropped due to its difficulty. Oberlin is not your run-of-the mill institution and is noted for its excellence in general education. An understandable argument can be made that subject matter expertise, or at least pedagogical expertise, is uncommon to arts educators across grade levels in a single discipline, a topic I address later.

In this *mélange* termed arts education, including each individual discipline, there are presently, despite ten years of standards, no identifiable common outcomes such as a skill, process, product, or knowledge. There is certainly no corpus of music, art works, plays, or dances (or creators and performers) in each of these fields that would be familiar to students graduating from high school who have participated in present experiences.

Contributing to reader madness, I have digressed from definitions of evaluation to describing the context for any definition. Let me now distinguish the boundary between assessment and accountability. It is accountability that the public wants from the schools. The public wants to know how priorities are established, they want data on the success of those priorities, and they want the consequences (rewards and punishments) for students and teachers to relate to the schools' priorities. The accountability movement was energized by the reports on how little high school students know about geography and American history. Recognition of the lack of historical knowledge led to an investigation of the content of social studies courses, in which critics found an emphasis on multicultural education and an integration of subjects! Marshalling data, they informed the public that ethnic groups in the U.S. grew from 1500 identifiable congeries in 1990 to 5000 in 1996 (Rochester, 2003, 28)! If cultures are to become a curricular subject, each must be well taught so valid comparisons among cultures can be made. Selecting the cultures from among 5000 without offending those not selected would challenge the best teachers. In the same study, integration of visual art with social studies was criticized because there has been a seemingly random selection of topics, topics

that emphasize the superficial or exotic such as clothing styles, food, holiday, religious observances, leisure activities, rituals, and other customs (Rochester, 2003, 45). Rochester states: "More often than not, such features are stressed mainly to provide a sense of difference and to 'celebrate diversity' without much context to give them real meaning. To understand a culture, the curriculum must be designed to explain linkages among family structure, kinship grouping, language, technology, religion, art, and ethnical norms and laws" (45). Having students tie-dye textiles to integrate art and ethnicity in social studies using modern-day cloth and nontoxic commercial dyes is cited to demonstrate inappropriate experiences (51). Leming states that of the 63 articles published in *Social Studies Theory and Research in Social Education* between 1992 and 1997, none examined the influence of social studies curriculum on student acquisition of historical or civic knowledge (Leming, 2003, 136). This is one example from the hundreds of "horror stories" about contemporary education. Accountability is about clarifying the boundaries within the educational establishment for better understanding of responsibility. Linda Crocker is correct that "the tsunami of educational accountability is at our door" (Crocker, 2003, 10).

Schwandt suggests that *accountability* is a technical and contractual notion; *responsibility* in education is a moral notion (Schwandt, 2003, 362). What we want in the arts is responsibility. As an aside, I enjoyed a recent ruling by a North Carolina judge that the state is responsible (accountable) for poor student performance -- not the students, teachers, or schools. When education was truly a local enterprise, the schools and teachers were accountable to the local school board and to the community. The data used to determine accountability was informal. The competence of graduates was common knowledge drawn from their performance in local employment or in higher education. As communities grew into cities and the school committee became responsible for a large number of elementary and secondary schools, accountability required formal measures; lacking these, courses such as art and music could be ignored. Music contests and public displays of art works filled the need to demonstrate their value to the community. Horace Mann is often credited (or blamed) for initiating formal accountability. He was the first state superintendent of instruction (the state was Massachusetts), and he took his position seriously. In visiting Boston schools, he found wretched instruction;

students displayed only superficial knowledge at the annual PTA meeting but with no understanding. Mann ruled that, as a consequence, Boston was to receive no further support from the state unless the teaching was improved (Parsons, Howe, and Neale, 1845). Mann's ideas for assessment were put into place in 1845 and 1846, but abandoned in 1847 because the results were never used (Black and Kline, 2002, 224). In the more than 150 years since Horace Mann, school districts have grown even larger, the state has assumed a greater responsibility for financial support, and the distance between the student and a state legislator is now sufficiently great that the original understanding of accountability has been lost. Today, a five to seven year cycle of assessing a school might be feasible but more frequent assessment is not necessary. Teachers don't change objectives and materials on an annual basis. In 2004, in addition to the state's messing with student learning, there is a federal role in education: the federal government funds around eight percent of the costs, and though this is a small percentage it represents a sizeable amount of money, sufficient to impose federal legislation, most recently No Child Left Behind.

THE REFORM MOVEMENT AND THE CURRICULUM

Reform

There are many individuals and groups who wish to "help" establish the school's priorities and here everything is in play: policy, politics, editorials, set pieces, business, random research reports, books, professional organizations, and journal articles. The reform movement may have been initiated over a concern for student learning and the welfare of kids but today the overriding concern is control of the schools, primarily through controlling the curriculum and testing. Charter schools, vouchers, and home schooling relate to control as well as the approximately 100 extant comprehensive school reform plans among which are the well-known Success for All, Achieve, Comer School Development Program, Coalition of Essential Schools, Modern Red School House, Core Knowledge, Roots and Wings, Audrey Cohen College System of Education, and Achieve. Also wishing to influence the curriculum are advanced placement, state departments of education and accrediting agencies. Data from

evaluations are accumulating on each of these, documenting their relative effectiveness, and it may be important to know the priority of the arts in each program and for each player.

Standards

National standards are the anchors in the reform movement with advocates for change (or control) arguing from published standards. The standards have been challenged in many subjects and subsequently revised, but the standards in the arts have not even been seriously debated. One guess is because those standards are both broad and vague, do little more than describe desired experiences, and can be ignored. The arts standards are not standards; rather they provide broad aims from which teachers (and students) will, it is assumed, derive instructional objectives and the appropriate standards for these objectives. The idea of having standards is popular with the public and 49 states have established content standards in “core” subjects. Core is one of the fuzzies ---it always means at least math and language arts and may include science. Thirty states hold schools accountable using test scores and 23 states can impose sanctions on low performing schools. Nineteen states require exit or end of course examinations for graduation with five additional states preparing such examinations (Crocker, 2003, 6). Although the arts standards are primarily nominal, they do serve to constrict many traditional popular arts experiences in schools. Where the teacher is expected to “cover” all of the standards, the result is that experiences are superficial and produce little understanding. The madness with arts standards occurs because they are promoted at a time when opportunity to learn has been reduced not only for budgetary reasons which may be temporary, but due to increased weight on other subjects and a move to ensure that unwilling and disadvantaged students also become proficient in two or three core subjects. Evaluation measures presently suggested to accompany the voluntary national standards in the arts lack substance and may even be harmful to long-term learning. One concern is that the curriculum and evaluation must be aligned. Of greater importance is alignment of the curriculum and the standards.

Curriculum

Experience and observation indicate that there are two distinct

curricula in music and in drama, one in “required” education and a separate one for “elective” education, each “requiring” its own standards. Without clarity about the curriculum, the fuzzy border of what content belongs in the discipline of music or theatre will prevent any valid progress. The division between required and elective goals is not one of grade level although in practice this tends to be the case. Teaching required music and theatre, at any level, requires competencies not required in elective arts. In music, for example, the subject matter expertise that is expected today of secondary school teachers will require the secondary music teacher to know, in some detail, each of the band and orchestra instruments, guitar, piano, and voice, as well as how to teach individually and in group situations. The curricular distinction becomes clearer when one analyzes teacher education in music where the requirement for high-level musical skills is generalized to all teaching situations; a requirement that contains a message for alternative certification programs in the arts.

The New York Times on October 9th, 2003, interviewed Stephanie Blyth, a star mezzo soprano in the opera world who obtained a music education degree at the Crane School of Music in Potsdam, New York, one of the better institutions for music education in the United States. Not until her practicing teaching experience in elementary school music did she realize that she was ill-equipped and unsuited for success in that field. (She reports that for the following semester she majored in marijuana, [Smith, 2003, 5].) Ms. Blythe had enjoyed singing in her high school chorus but the issues of elementary music are so dissimilar from high school chorus, issues that did not surface in her four years of teacher education, that she began student teaching totally unprepared. Applying this scenario to evaluation, it takes little imagination to believe that assessment techniques for second grade children are not a priority in the music teacher education curriculum. Theatre is almost as diverse. Theatre can be creative drama, language arts, or any subject where language, literature, and their use is emphasized. Creative drama is usually taught in elementary and middle schools with many non-arts objectives. The objectives for understanding theatre, its history, literature, and production are primarily found at the high school level. The dual curriculum is not as noticeable in visual arts because production has been the primary focus of all visual arts education until the introduction of Discipline Based Art Education with its addition of history, analysis/appreciation,

and aesthetics to the arts curriculum. DBAE did broaden the curriculum in many schools although it is not found as a K-12 “program” in most schools. In high school, production reigns supreme. The teaching of dance is so rudimentary that it is difficult to surmise whether the public would be accepting of it as an art in the definition of a well-rounded education.

Given their history, the fuzzy boundaries and fuzzy content of courses in the four art forms is not surprising. Visual arts and music both began in the Boston Public Schools early in the 19th century. Visual arts preceded music because the ability to draw accurately was a necessity in building America in the industrial age. Music was justified as a required subject because of its health benefits and the hidden good of improving congregational singing in Boston’s Protestant churches. Secondary arts have been elective and often extra-curricular. Academic credit for these courses was added more for bureaucratic control than due to a belief that the outcomes were equal to those in trigonometry or English literature. Whether grades in these secondary arts courses should count in a student’s grade point average for graduation or college admission remains controversial in 2004. The secondary arts performance programs are not dependent upon outcomes from K-6 instruction and there is limited content commonality in their curricula across schools or even within a single district. The flexibility that an arts teacher has in making curriculum decisions is unmatched and may explain why arts teachers in American high schools are reluctant to adopt all of the standards and any imposed evaluation based upon these standards. This flexibility is not limited to the secondary schools. In the college curriculum for the prospective arts teacher, there is little commonality in method courses either in objectives, content, or experiences, and only slight commonality in course work on the philosophy of arts education. We have no data in any of the arts that indicates the strength and weaknesses of any of the methods taught nor do we have evidence, other than the ability to model, that the performance experiences of the teacher contribute strongly to student outcomes. Flexibility becomes madness in deciding what to evaluate. Weeden argues that assessment remains the weakest aspect of teaching in most subjects (Weeden, 2002, 41). Only a small leap of faith is needed to be confident that the situation he describes is no better in arts teacher education.

Research by Achinstein working with 37 experienced teacher induction leaders confirmed that new teachers did not know much about assessment; only 35% could align curriculum and standards, 24% knew about reflection and 38% knew how to use assessment to guide their own growth (Achinstein, 2003, 1496). In a charter school in California experimenting with differential salary options for teachers, professional development is based entirely on competence in the visual arts. (In *No Child Left Behind*, the definition of which art(s) to include as a core subject is the responsibility of each state.) To obtain a salary bonus, the arts teacher is to be exemplary in use of traditional art forms (e.g. drawing, painting, collage, design, and exploration in media arts), must promote the use of art forms in other subject areas, use appropriate materials and teaching strategies, implement appropriate student activities, and consistently plan cooperative group projects and individual production to fully engage all students actively (Kellor, 2003, 66). What is expected of arts teachers in California elementary schools borders on the extreme fuzzy, or extreme madness.

No Child Left Behind

The federal legislation that has become known as NCLB (No Child Left Behind Act of 2001, Public Law Number 107-110, 2002) is both a political and an accountability document. The intent to have all students at a “proficient” level in core subjects (language arts, math, and science) by 2014 is an admirable goal. Politically, its intention was not only to demonstrate President Bush’s commitment to education but to give school districts and states political “cover” from any fallout from the extensive testing required. Rewards and sanctions come as a result of tests and of annual yearly progress toward student proficiency. The required evaluation is to be reported not only by school, but district, state, region, and nation. Within the school, identifiable groups (e.g. race, SES, gender, and potential) will also be judged. The specific assessments are not mandated and can consist of locally constructed tests, homework assignments, portfolios, interviews, observations, projects, and presentations (Bhola, 2003, 21). The only requirement is that these assessments measure the knowledge and skills deemed valuable and described in policy documents at the local level. It seems obvious that these documents will be state or national, rather than local. The National Assessment of

Educational Progress exam is to be administered periodically and Gerald Bracey argues that this requirement will mean that NAEP's definition of proficient will be the standard, not local or state definitions (Bracey, 2003, 149). The levels established by NAEP in core subjects, however, have been found flawed by any number of groups, ranging from the General Accounting Office, to CRESST(National Center for Research on Evaluation, Standards, and Student Testing), and the National Academies of Education and Science.

Robert Linn, in his 2003 AERA presidential address, suggests that if progress continues to be made at the same rate as in the past decade, it will be 2056 before 4th grade students will be proficient in math, 2060 for 8th graders and 2166 for high school seniors (Linn, 2003, Chicago, April 23)! There are numerous technical problems inherent in such a massive evaluation effort, and costs to school districts will go well beyond that envisioned by the legislation.

The levels proposed for the arts are even more arbitrary and have not been subjected to any analysis. They were established by a committee that communicated by mail and were intended to initiate a discussion in the profession on performance standards. Although performance standards are the standards of most importance, attention has been focused exclusively on content standards. It is nearly inconceivable that the arts will be tested, although advocates, recognizing that subjects to be tested are subjects that are taught, will suggest the importance of assessment and accountability in the arts and will continue to promote at least an NAEP examination in the arts. The arts do not have "programs" like core subjects; hence program evaluation in any art form is inappropriate except as a case study in a single school. As an arts program is pretty much whatever the teacher decides, great flexibility has been given to arts advocates as they can tailor their claims -- to individual audiences, to outcomes resulting from both in-school or out-of-school instruction, to either arts outcomes or outcomes of character and diligence. NCLB does create a soapbox from which to preach that the curriculum that omits the arts is too narrow, that education in the arts is presently equally unsatisfactory, and that educational balance is critical for full and enlightened participation in American democracy. (In too many cases, however, the public and many arts teachers are satisfied with the status quo.) NCLB also creates an argument for arts specialists. As with the pressures of NCLB testing, the classroom teacher would have to forfeit any time she

has had for arts instruction.

Classroom teachers are wailing about too much testing or testing of the wrong kinds, some of which wailing is prompted by possible consequences for the teacher or the school when comparisons are unfavorable. Neither students nor parents are opposed to today's testing mentality as indicated by the PDK/Gallup Poll (2002) and interviews with students. Sixty-seven percent support annual testing of all students in grades 3-8 and 68 percent favor use of a single national test (PDK 2002 poll). Only 30 percent believe that there is too much testing in the schools. High stakes tests are acceptable to those polled; (PDK/Gallup Poll 2003, 45) the concern is focused more on the use of a single test as the basis for decision-making. Arts educators are caught in the increasing madness over the appropriate role for external exams in student achievement. Kim Marshall, a Boston elementary school principal, reports that only the adoption of the high-stakes state test (MCAS) brought positive changes to his school (Marshall, 2003, 105-113). Similarly, Edward Humes, in *School of Dreams* also reports on the necessity of competition and testing to "make the grade" in a high school that challenges students to their full potential (Humes, 2003). In the midst of all of this, arts educators are pleading for more ambiguity in student lives. But ambiguity is a quality that American society seeks to avoid, bombarded as it is by continuous news reports from all corners of the planet and outer space that shake our thought patterns and assault our stability. The public is not likely to see ambiguity as a critical outcome, even that ambiguity experienced through the arts (Tineke, 2003, 288).

College

College teachers often model the behavior expected of their graduates—the student works in the chemistry laboratory alongside a professional chemist; the drama teacher directs plays; the applied music teacher performs in public; and the teacher educator is expected to be an inspiration in the classroom. Unfortunately, teaching expertise does not distinguish, the faculty in the college of education, because excellent teaching is found across the campus.

Pertinent to this article is the absence of evaluation in American colleges. Evaluation essentially began at the college level in 1985 (Banta 2002, p 1). Of course, college programs have been approved by professional organizations and professors have given mid-semester

and final exams, but college administrators have not employed any mechanism to determine what students have learned as a result of their college majors, how any learning compares with what should have been learned, or what is being learned at comparable institutions. The primary concern has been with quantitative data on drop-outs, transfers, and job placement. Based upon a 1998 survey and returns from 1393 institutions, 78 percent of those institutions admitted to giving no attention to learning outcomes (Peterson and Vaughan, 2002, 31). Preparing for accreditation (69%) was the most important reason for engaging in student assessment (33). Public schools may be overwhelmed with assessment requirements but they lead in experience and knowledge when compared to evaluation in higher education. College administrators are like their colleagues in the public schools in claiming that generic critical thinking and problem solving skills across the curriculum are the important objectives and the reason for general education (Erwin and Wise, 2002, 69). It is no wonder that the “me-too” arts educators and arts advocates include problem solving as an outcome of high school arts even though that objective has never occurred to enrolled students or the arts teacher.

Assessment

A reasonably thorough analysis of educational issues in 2004, with special attention to the extensive clamor about assessment results is that summative assessment is important in the arts but not a high priority. Contests, public performances and exhibitions are an important element in arts education and serve a comparable purpose to the Cincinnati Symphony Orchestra’s annual tour to New York City and the accompanying New York Times review. Performance standards in each of the arts are being “globalized” and individuals and groups are compared across international boundaries, a comparison that relates to maintaining standards. Competition and comparisons in the schools can also be educationally beneficial when approached properly. When a second grade child cries because of some type of evaluation process, the evaluation has been ineptly presented (Crocker, 2003, 10).

Formative Evaluation

The primary concern of every teacher is to provide feedback to

students on daily and weekly objectives. The teacher must also keep records. Without documentation even on oral feedback, the teacher (and student) has limited data upon which to individualize instruction and develop improvement strategies. Feedback consists of in-class questions and comments as well as group measures that include written exercises, projects, performances, auditions, and more. Every kind of assessment can be valuable to teaching and learning. A major drawback is that evaluation requires time, time that is already in short supply. The point of formative assessment, the kind the teacher employs regularly in the classroom, is to make instruction more efficacious. If the assessment is employed as it should be, the improved learning will more than make-up for the “lost” time.

Teachers learn about helpful assessment strategies through coursework, professional development, and experience. Unfortunately, our knowledge of the role of learning styles, and even how students learn, remains limited. Nevertheless, this paper, will conclude with a few ideas that have proven successful that reflect what we tentatively know about teaching and learning. First, almost no empirical research exists to support the idea that authentic assessment, self-assessment, or peer assessment is more intrinsically motivating to students than traditional measures (Erwin and Wise, 2002, 70). What we do know, counter to what is published in arts journals, is that students whose motivation in required courses is already low are adversely effected by more challenging assessment tasks. For these students, multiple choice tests are the better option (Wolf, Smith, Birnbaum, 1995, 341-351).

Assessment is not a new topic in the arts; one model is the individual music lesson where the student receives a clear assignment of what is to be accomplished and demonstrates the results of practicing at the next lesson. At this lesson, immediate feedback is provided whenever the teacher believes that an improvement is possible. Often this feedback is accompanied by modeling on the part of the teacher. A similar model is found in theatre and dance. Similarly, visual arts at both the elementary and secondary levels, when production is an objective, is characterized by individual assessment and immediate feedback,---the peripatetic teacher providing feedback on the process and the product. This combination of instruction-assessment continues into the out-of-school situation in the arts and sports, where professionals employ coaches whose primary task is to

identify anything that might interfere with exemplary performance. Rehearsals with dance, theatre, and music groups—including at the Metropolitan Opera—are marked by a well-integrated assessment process, in these cases evaluation dictating instruction. With amateurs, the order is likely reversed: instruction comes first, with the student involved in trial and error and then receiving feedback. Some coaches, instructors, and directors are more adept than others -- indicating the need for both instruction and experience in assessment procedures. Teacher education more often than not fails to provide opportunities and instruction in (1) identifying a general learning problem, (2) selecting the individual(s) where the problem centers, (3) identifying the specific need, and (4) suggesting an appropriate solution --- all based upon student level. Having a single solution to a problem is a first step, having three or more feasible solutions to an instructional issue is learned from instruction and experience. Most problems that learners have in the arts have been previously noted and a range of solutions exist -- one does not need to learn to assess exclusively “on the job.” Unfortunately, the task of assessment is more difficult with young students learning in groups, exacerbated by heterogeneous grouping where a full repertoire of problem identification and solutions needs to be “on-call.”

Fair assessments increase student self-esteem because self-esteem is earned, not given by the teacher. Beran suggests that today’s public school systems shrink from giving students the constant challenge required to move to higher levels of mastery and insight. He believes that “accommodating” to inner city kids (or anyone disadvantaged) results in a loss of self-esteem and respect. “The dumbing down of the curriculum, the unwillingness to make kids learn a body of knowledge and develop basic skills through drill, the easy tests, and lack of consequences for leaving homework undone --all conspire to keep kids’ horizons low, instead of expanding them” (Beran, 2003, 25). As arts teachers know, drill is required in skill development; practicing and rehearsing is our equivalent to homework. A strong predictor for excellence in any art is not a test but knowledge whether the student practices that art at home.

Self-evaluation is to be encouraged but, again, it is overrated. There would be less need for coaches and directors (with professional artists) if self-evaluation were easy and did not interfere with further learning. Hewitt, working in music, found self-evaluation effective only

for improving intonation (Hewitt, 2002, 226). With self-assessment, students tend to focus on how hard they worked and how far they have come. If the student is initially deficient, a difficulty arises between providing a reward for progress and challenging the student so that the deficiency does not continue.

Evaluation in the arts is far more complex than in other subjects due to instruction out of school and the influence of the student's immediate and peer culture. The task is likely to become more complex if the present fuzzy definitions of arts education are expanded further. Visual arts has, for some time, included museum education; with 16,000 museums in the US, and the number of visitors to museums exceeding that of all sporting events combined, separating in and out-of-school learning is complex (Paris, 2002, 38). We know from the research of Anderson (1987) and Hein and Alexander (1998) that students do learn from museum visits. Kerry Freedman has recently published a text advocating that the boundaries of visual arts be expanded even further to include all visual stimuli that one encounters including ads, TV, and graffiti. Talk about controlled madness, at least the responsibility should be shared with humanities!

Any use of evaluation will find resistance among some arts educators who desire the status quo and do not want to risk failure. These individuals have an established power arrangement within the school and any change in resources is a threat. They prefer vague objectives and vague assessments (Taut and Brauns, 2003, 255).

Assessment has more impact on learning when the objectives are clear to both student and teacher. Thus, it is important to establish who defines the norms of the discipline and the criteria for setting performance levels. The voluntary national standards do not provide this clarity, and the professional organizations have not initiated the discussion necessary for setting norms. Lacking norms, outcomes presently differ vastly from school to school. Present arguments over what is quality instructional material emphasize inputs over outputs, thus bypassing emphasis on competencies as outcomes. Research data on causal relationships between instruction and learning do not exist; still students need to know some ways of how to determine what to value, what to attend to, and how to use perception in understanding the arts.

The teacher needs to understand that students come to arts classes with ideas about the relationship between talent and effort.

Today's population is about evenly divided on the importance of each. The youngest students attribute success more to effort than do students above the age of nine or 10. Dweck classifies students as having (1) an entity or fixed theory about ability or (2) basing their learning on an incremental theory, by which she means that ability is malleable. (Dweck, 1999, 20). Weeden, Winger, and Broadfoot (2002) use the term helpless and mastery children. Helpless students are those who are motivated by the desire to be seen to do well; accept the idea that if they fail it is because they are not clever enough; if a task is difficult, there is nothing they can do about it; and they avoid challenges. Feedback from any assessment must differ according to these two student types. Helpless students are hindered by undeserved praise. For group and individual endeavors, mastery students benefit from last-minute emergency meetings when, collectively, the questions are asked: How well are we doing? How well should we be doing at this stage? What must be done to make the performance or exhibition a success?

Assessment criteria must be in language understood by the students and language that indicates the fairness of any assessment. A teacher can prepare for an assessment but has little control over the results, and it is the results that have the most impact on learning. Interpreting the results for students, parents, and school administrators is complex, but this is an area where instruction is available, and it must become part of professional development and teacher education in the arts. It is known that youngsters at the age of five who study piano learn to read music notation in two clefs and with some understanding, yet there are students who have completed required music in elementary school and are unable to make sense of musical notation. Teachers may not understand, nor have been given an education that informs them, what is an appropriate challenge.

Before assessment can be helpful, the curriculum must be aligned with expected standards and with the projected assessment. Core subjects have research data on various strategies for aligning, e.g. Survey of Enacted Curriculum and Council for Basic Education (Bhola, Impara, Buckendahl, 2003, 22.). Where no curriculum exists or is followed, however, the arts are unready to employ the complex assessment strategies found in education journals. For example, in visual arts, Charles Dorn attempted to teach teachers in 50 states how to use rubrics to judge art works. In addition to the rubrics, he reports

that the teachers used their intuitive knowledge in arriving at reliabilities of .345 and .442 after training. Such low reliability indicates the difficulty of arts assessment. Rubrics are best used in summative evaluation, and they need to be established (descriptions written) after competent judges have evaluated work and placed the works in the suggested number of categories. Rubrics are also not generalizable, applying primarily to one population at a time.

Teachers need to be told, and often, that the focus of arts assessment is in the classroom, by the teacher, and that the external assessments produced by today's testing madness do not apply. State tests and the national assessment are not models for classroom testing, but just the opposite, for they do not provide immediate feedback, are not written in student language, and are not aligned with the instruction that has been conducted. Most arts classes have few routines and limited stability and might require frequent changes in assessment strategies. Authentic assessment, in any field, performing a task only once, has limited reliability and a large measurement error and is not, by itself, a valid indicator of what a student has learned or can do. Average or mean scores, such as those published by the state, lead only to acceptance of mediocrity in the arts. Arts students invariably have as their model one who excels in the art form. Students participating in contests do not compare themselves with the average attendee---they want to be compared with the best.

Portfolios cannot be compared within a class or against any standard and hence have limited use. Colleges that have tried them have found them ineffective measures at the end of a course with possibly more value where they can be continued over several semesters and several courses (Palomba, 2002, 210). Observation does not tell us why things do or do not happen, thus, its diagnostic value is limited although it remains important for instruction and other roles in assessment. We need definitions not only of competence but also of incompetent students and teachers if assessment is to fulfill its potential.

Much more can be said; every assessment tool has potential and limitations. These remarks offer only a partial listing of what needs to be considered in a course in evaluation for all teachers in the arts. Evaluation requires understanding and time; when evaluation is employed as a Shock and Awe experience rather than part of daily learning, more negative than positive learning will take place.

SUGGESTED CITATION

Colwell, R. (2004). Evaluation in the arts is sheer madness. *ArtsPraxis*, 1 (1), 1-24.

REFERENCES

- Anderson, D. (1987). *A commonwealth: Museums and learning in the United Kingdom*. London: Department of National Heritage.
- Angelo, Thomas A (2002). Engaging and supporting faculty in the scholarship of assessment: Guidelines from research and best practice. In Banta, Judy W. (Ed.) *Building a Scholarship of Assessment* (pp.185-200). San Francisco: Jossey-Bass.
- Athanases, Steven Z. and Achinstein, Betty (2003). Focusing new teachers on individual and low performing students: The centrality of formative assessment in the mentor's repertoire of practice. *Teachers College Record*, 105 (8), 1486-1520.
- Banta, Judy W. and Associates (2002). *Building a scholarship of assessment*. San Francisco: Jossey Bass, 1.
- Beran, Michael Knox (2003). Conservative compassion versus liberal pity. *City Journal*, 13 (2) 16-27.
- Black, Karen E. and Kline, Kimberly, A. (2002) in Banta, Judy W. and Associates. *Building a Scholarship Of Assessment* (p.223-239). San Francisco: Jossey-Bass.
- Bhola, Dennis S., Impara, James C, and Buckendahl, Chad W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22 (3) 21-29.
- Borko, Hilda, Wolf, Shelby A., Simone, Genet, and Uchiyama, Kay (2003). Schools in transition: Reform efforts and school capacity in Washington state. *Educational Evaluation and Policy Analysis*, 25 (2) 171-201.
- Bracey, Gerald D. (2003). We're number one (again), *Phi Delta Kappan*, 85 (1) September, 87-89.
- Bracey, Gerald D (2003). The 13th Bracey report on the condition of public education. *Phi Delta Kappan*, 85 (2) October, 148-164.
- Burack, Jonathan (2003). The student, the world, and the global education ideology. In Leming, James, Ellington, Lucien, and Porter, Kathleen, (Eds.), *Where did social studies go wrong?* (pp.

- 40-69). Washington: Thomas Fordham Foundation.
- Cohen, David K., Raudenbush, Stephen W., Gall, Deborah (2003). Resources, instruction and research. *Educational Evaluation and Policy Analysis*, 25 (2), 119-142.
- Consortium of National Arts Education Associations (1994). Dance, music, theatre, visual arts: What every young American should know and be able to do in the arts. Reston, VA: Music Educators National Conference.
- Crocker, Linda (2003). Teaching for the test: Validity, fairness, and moral action. *Educational Measurement: Issues and Practice*, 22 (3), 5-11.
- Davis, Jessica (2003). In defense of failure: A rule for the arts in non-arts education. *Education Week*, Volume XXIII, Number 6, October 6, pp 28, 30.
- Dorn, Charles M. (2003). Models for assessing art performance (MAAP): A k-12 project. *Studies in Art Education*, 44 (4), 350-371.
- Dweck, Carol S. (1999). *Self-Theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.
- Education Week* (2003). Critics question federal funding of teacher test. Author. *Education Week*, Volume XXIII (6) October 8, 1, 7.
- Erwin, T. Dary and Wise, Steven L. (2002). A scholar-practitioner model for assessment, in Banta, Judy W., (Ed.) *Building a scholarship of assessment* (pp. 67-81). San Francisco: Jossey-Bass.
- Freedman, Kerry (2003). *Teaching visual culture: Aesthetic and the social life of art*. New York: Teachers College Press.
- Fishkin, James (2003). Informed public opinion about foreign policy: The uses of deliberative polling. *Brookings Review*. Summer, 16-19.
- Hamilton, Laura S. McCaffrey, Daniel F. Stecher, Brian, M. Klein, Stephen P. Robyn, Abby, and Bugliari, Delia (2003). Some large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25 (1), 1-29.
- Hatch, Thomas (2001). Incoherence in the system: Three perspectives on the implementation of multiple Initiatives in one district. *American Journal of Education*, 109 (4), 407-437.
- Hein, G.E. and Alexander, M (1998). *Museums: Places of learning*.

- Washington, D.C.: American Association of Museums.
- Hewitt, Michael (2002). Self-evaluation tendencies of junior high instrumentalists, *Journal of Research in Music Education*, 50 (3), 215-226.
- Humes, Edward (2003). *School of dreams: Making the grade at a top American high school*. Orlando: Harcourt, Inc.
- Instrumentalist* (2003) Survey of school music budgets. *Instrumentalist*, 58 (1), August 20-29.
- Kellor, Eileen, M. (2003). *Catching up with the Vaughn express: Four years of performance pay and standards based teacher evaluation*. Wisconsin Center for Education Research, Working Paper Series 03-02.
- Kuh, George D, Gonyea, Robert M. and Rodriguez, Daisy P. (2002). The scholarly assessment of student assessment. In Banta, Judy (Ed.), *Building a scholarship of assessment*, (pp. 100-127). San Francisco: Jossey-Bass.
- Leming, James S. (2003). Ignorant activists: Social change, "higher order thinking," and the failure of social studies in Leming James, Ellington, Lucien and Porter, Kathleen, (Eds.) *Where did social studies go wrong?* (pp. 124-142). Washington: Thomas Fordham Foundation.
- Leming, James, Ellington, Lucien and Porter, Kathleen (2003). *Where did social studies go wrong?* Washington: Thomas Fordham Foundation.
- Linn, Robert L (2003). Accountability: Responsibility and reasonable expectations. Presidential Address To American Education Research Council. *Educational Researcher*, 32 (7), 3-13.
- Linn, Robert, Baker, E, and Betebenner, D (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31 (6), 3-16.
- Mabry, Linda (1999). Writing to the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan*, 80 (9), 672-679.
- Manzo, Kathleen (2003). Teachers picking up tools to map instructional practices. *Education Week*, Volume XXIII (6) October 8, 8.
- Marshall, Kim (2003). A principal looks back: Standards matter. *Phi Delta Kappan*, 85 (2), 105-113.
- Miller, G. Edward (2003). Analyzing the minority gap in achievement

- scores: Issues for states and federal government. *Educational Measurement: Issues and Practice*, 22 (3), 30-36.
- No child left behind act of 2001 (2002). Public Law Number 107-110.
- Palomba, Catherine A. (2002). Scholarly assessment of student learning in the major and general education. In Banta, Judy W. (ed.) *Building a scholarship of assessment* (pp. 201-222). San Francisco: Jossey-Bass.
- Paris, Scott G (2002). *Perspectives on object-center learning in museums*. Mahway, NJ: Erlbaum.
- Parson, T. Howe, S.G., and Neale, R.H. (1845). 1845 report of the annual examining committee of the Boston grammar and writing schools. *The Common School Journal*, 8, 27-30.
- Peterson, Marvin W. and Vaughan, Derek S. (2002) Organizational and administrative dynamics that Support student assessment. In Banta, Trudy (Ed.), (2002) *Building a scholarship of assessment*. (pp. 26-48). San Francisco: Jossey-Bass.
- Pike, Gary R. (2002). Measurement issues in outcomes assessment in Banta, Judy W. (Ed.), *Building a Scholarship of assessment*. (pp. 131-147). San Francisco: Jossey-Bass.
- Rochester, J. Martin (2003). The training of idiots: Civics education in America's schools. In Leming, James, Ellington, Lucien, and Porter, Kathleen (Eds.) *Where did social studies go wrong?* (pp. 6-39). Washington: Thomas Fordham Foundation.
- Rose, Lowell and Gallup, Alex. (2002) The 34th annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 84 (1), September, 41-56.
- Rose, Lowell, and Gallup, Alex (2003). The 35th Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 85 (1), 41-56.
- Schwandt, Thomas (2003). Back to the rough ground! Beyond theory to practice. *Evaluation*, 9 (3), 353-364.
- Sidse, Sverdrup (2003). Toward an evaluation of the effects of laws: Utilizing time-series data of complaints, *Evaluation*, 9 (3), 325-339.
- Silverman, D. (2000). *Doing qualitative research: A practical handbook*. London: Sage.
- Smith, Dinitia (2003). A mezzo poised for the heights, *The New York Times*, Volume CLIII (52631) October 9, B1.
- Taut, Sandy and Brauns, Dieter (2003). Resistance to evaluation: a

- psychological perspective, *Evaluation*, 9 (3), 247-264.
- Tinke, A. Abma, and Noordegraaf, Mirko, (2003). Public managers amidst ambiguity: Towards a typology of evaluative practices in public management. *Evaluation*, 9 (3), 285-306.
- Tomlinson, Carol Ann (2003). Deciding to teach them all. *Educational Leadership*, 61 (2), 6-11.
- Weeden, Paul, Winger, Jan, and Broadfoot, Patricia (2002). *Assessment: What's in it for schools?* London: Routledge/Falmer.
- Wolf, L. F., Smith, J.K., and Birnbaum, M. E. (1995). Consequence of performance, test motivation, and Mentally taxing items. *Applied Measurement in Education*, 8, 341-351.
- Wright, Barbara D. (2002). Accreditation and the scholarship of assessment. In Banta, Judy W. (Ed.), *Building a scholarship of assessment* (pp. 240-260). San Francisco: Jossey-Bass.

AUTHOR BIOGRAPHY

Richard Colwell, a Guggenheim and Fulbright scholar, is a member of MENC's Hall of Fame. He edited the *Handbook of Research on Music Teaching and Learning* and co-edited, with Carol Richardson, *The New Handbook of Research on Music Teaching and Learning*. He founded the *Bulletin of the Council of Research in Music Education* and the *Quarterly*. He has published tests with both Follett and Silver Burdett. He authored the arts section of ASCD's Curriculum Handbook, the arts section for Education Research Services as well as the music entries for the Encyclopedia of Education and the Groves and Harvard Dictionaries of Music. He has been on the faculties of Colorado, Illinois, Michigan, Georgia State, Boston, and the New England Conservatory of Music, of which, he was chair of music education at three and distinguished faculty member at the others.