Validity, Reliability, and Equity Issues in an Observational Talent Assessment Process in the Performing Arts

Barry A. Oreck, Steven V. Owen, & Susan M. Baum

The lack of valid, research-based methods to identify potential artistic talent hampers the inclusion of the arts in programs for the gifted and talented. The Talent Assessment Process in Dance, Music, and Theater (D/M/T TAP) was designed to identify potential performing arts talent in diverse populations, including bilingual and special education students and students who have had no prior formal arts instruction. Research results over 13 years in elementary schools in New York and Ohio provide evidence that creative and artistic potential can be assessed validly and equitably and that such assessments can offer a reliable prediction of success in rigorous arts instruction.

Introduction

The development of valid, equitable, performance-based assessments in the performing arts poses many challenges, but also offers a wealth of valuable techniques and approaches that can be widely applied in education. Valid arts assessments, whether of achievement or ability, must take into account a number of especially complex and potentially confounding issues that include the wide variety of psychological and educational constructs involved in art making, the difficulty of inspiring artistic behaviors under assessment conditions, and the extreme variation in students' prior experience and access to arts instruction. Further, cultural and stylistic differences within art forms threaten the external validity of assessments and their applicability to various settings and populations.

These challenges are not insurmountable, however. The structure of arts activities, engagement of students in active participation

---

Barry A. Oreck is Adjunct Professor at Long Island University, Brooklyn, NY. Steven V. Owen is Professor and Senior Biostatistician in the School of Nursing and the Department of Preventive Medicine and Community Health at the University of Texas Medical Branch, Galveston. Susan M. Baum is Professor at the College of New Rochelle in New Rochelle, NY.

Journal for the Education of the Gifted. Vol. 27, No. 1, 2003, pp. 62–94.

and group processes, and the variety of methods for expressing ideas and feelings provide capable models for authentic, performance-based, curriculum-embedded assessment. Accountability in the arts, as in all subject areas, requires the development and testing of processes that are authentic to the discipline, psychometrically sound, aligned with the curriculum standards, and equitable to all students.

Since 1991, ArtsConnection, a New York City arts-in-education organization, has developed and tested new processes for the assessment of potential talent in the performing arts. The Talent Assessment Processes in dance (DTAP), music (MTAP), and theater (TTAP) were developed with grants from the Report to the Jacob K. Javits Gifted and Talented Students Education Program of the U.S. Department of Education. These are multisession, multiobserver, observational assessment processes designed for use in normal, nonarts magnet, K–12 schools. The initial research (ArtsConnection, 1993, 1996; Baum, Owen, & Oreck, 1996, 1997; Kay & Subotnik, 1994) has been expanded to Ohio, where schools are now mandated to assess students in the performing and visual arts for gifted and talented designation (Project START ID, 2001–03; Ohio Department of Education, 2000). This paper presents findings from two separate studies of D/M/T TAP and discusses the issues involved in designing and administering valid, reliable, and equitable talent assessment in schools with diverse student populations and limited access to arts instruction.

Background

Despite the inclusion of the arts in the U.S. Department of Education's definition of gifted and talented (Marland, 1972) and as a core subject in the No Child Left Behind Act (U.S. Department of Education, 2001), few schools, districts, or states require artistic abilities to be assessed along with other aspects of intelligence or academic performance. Many educators across the country subscribe to broader views of potential, such as those described in Renzulli's (1978) three-ring conception of giftedness, Howard Gardner's (1983) theory of multiple intelligences, and Sternberg's (1988) triarchic theory of intelligence, but most assessments focus solely on a narrow range of verbal test-taking skills (Darling-Hammond, 1994; Richert, 1992). The absence of valid and reliable assessment processes in the arts and the decreasing availability of arts instruction (Abeel, Callahan, & Hunsaker, 1994; Love & Kipple, 1995; U.S. Department

of Education, 1995) often leads teachers and parents to overlook the creative and artistic abilities of their children.

Validity Problems in Existing Assessments

Assessment in the performing arts, in particular, suffers from a lack of systematic, research-based methods. Many schools use one-time, high-pressure audition formats, videotaped performances, written tests, or teacher nominations to select students for magnet arts schools or special arts or gifted and talented programs (Byrnes & Parke, 1982; Elam & Doughty, 1988; Saunders & Schmidt, 1979). Each of these methods suffers from validity and reliability problems that prompt some arts educators to reject assessment in the arts as unfair and inherently elitist. Arts assessments that rely on loosely defined or subjective criteria or that are confounded by differences in students' prior instruction, cultural background, English language ability, parent involvement, academic standing, or behavior in other classrooms reinforce the notion that it is not possible to assess the arts fairly and systematically.

Few published talent assessment methods exist in dance or theater arts. Music, on the other hand, has a long history of aptitude tests. Most of these tests are based on listening and aural discrimination and involve a verbal or written response (Gordon, 1965; Seashore, 1938; Webster, 1994). A complete review of these measures will not be presented here, but their general validity and reliability problems will be discussed with regard to language dependence, cultural bias, and lack of predictive validity (Boyle & Radocy, 1987; Haroutounian, 2002).

Cultural and Stylistic Specificity. Screening for a few specific characteristics, such as a good singing voice or a specific body type, is an efficient way to select students for the study of a particular artistic style or technique, but it has limited generalizability. The New Ballet School of the Eliot Feld Ballets/NY, for example, assesses more than 15,000 elementary-grade students each year to select fewer than 350 for ballet training. In a short series of ballet exercises, the assessors are able to focus on a few obvious characteristics, such as a student's body type, flexibility, enthusiasm, and the ability to imitate what he or she sees (E. Feld, personal communication, 1998; K. Moore, personal communication, February 16, 1991). These ballet-specific criteria are the norm in many dance screening processes and have become codified in some published assessment tools. The Detroit Public Schools

Creative Products Scale (Byrnes & Parke, 1982) used for admission to a magnet arts high school includes such items as "length of legs and arms should balance body, long hips should not be too close to rib cage, toes approximately even, lifted arch-high instep, no pronation, wide metatarsal, well-proportioned legs" (p. 6). Plainly, the attractiveness of these characteristics is culture-based; a dissimilar list of characteristics would be defined in different cultures. The idealized picture of the dancer, musician, or actor carries pervasive cultural stereotypes that are codified by many style-specific assessments.

Language and Cultural Dependence. Many of the published talent-identification systems that use a written-response form are highly correlated with measures of academic achievement and verbal ability (Rainbow, 1965). One of the best known measures of musical aptitude, the Gordon Test of Music Audiation (Gordon, 1979), asks students to fill out a written-test form within 110 minutes. Some questions use musical terminology that may be unknown to students. After listening to two short musical phrases, the question is asked, "Are the two passages played in the same meter or different meter?" Students who have no formal musical training are unlikely to know the meaning of the term *meter*. The Gordon test also asks for value judgments that are based on culture-specific musical experiences. For example, a question on musical sensitivity asks, "The two passages are the same except for the ending; which ending is more appropriate?" The conventions of Western music, which comprise the music samples, suggest answers that likely differ in other musical forms and cultures.

In theater auditions, language tends to be emphasized in verbal, rather than written forms. While speaking is undoubtedly important in most aspects of theater arts, the common audition practice of reciting a memorized passage can reflect a reading as much as a speaking skill. This format is an obvious hindrance for many students, especially those who speak and read English as a second language. Dance, which seems like the least language dependent of the arts, also frequently includes written criteria for talent identification. The screening instrument developed by the Educational Center for the Arts in New Haven, Connecticut, for example, asks students to "list four body facings in ballet terminology, list four pioneers of modern dance, list four ballet centers, and list the live dance performances you have seen" (Saunders & Schmidt, 1979, p. 21). These questions require prior dance experience and reveal a clear stylistic bias.

Lack of Predictive Validity. Although some talent-identification systems have undergone extensive testing, little follow-up research exists to correlate identified potential with later achievement in the art form. Writers and researchers looking back at the early development of outstanding artists tend to focus more on motivational and environmental issues than on innate talent, except in studies of child prodigies (Bloom 1985; Subotnik 1995). Single-session assessments are vulnerable to fluctuations in testing conditions, students' moods and motivation, and other mitigating factors. Perseverance, practice, and deepening of understanding—all critical to successful talent development—are essentially disregarded in one-session assessments (Haroutounian, 2002).

Some gifted programs use tests of creativity or physical ability that are not specific to the art form, but are thought to bear a relation to artistic talent. Widely used tests, such as the Torrance Tests of Creative Thinking (Torrance, 1966), the Purdue Perceptual Motor Survey (Roach & Kephart, 1966), and the Basic Motor Ability Test (Arnheim & Sinclair, 1974), measure certain components that may be related to artistic talent, but have not been directly related to artistic potential or future achievement and cannot be considered valid predictors of success through arts training (Shwedel & Stoneburner, 1983).

Nomination and Screening Processes. Many schools use the recommendations of classroom teachers, parents, or arts teachers as nominations or prescreening for assessment or, in some cases, for gifted and talented identification. There are a number of research-based teacher recommendation instruments for gifted programs, but only a few include the arts. The most commonly used of these, the Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli, Smith, White, Callahan, & Hartman, 1976) and the Gifted and Talented Evaluation Scale (GATES; Gilliam, Carpenter, & Christensen, 1996), are completed by classroom teachers or arts specialists as one part of a talent-identification process. SRBCSS includes separate sections concerning musical, dramatic, and visual art behaviors and characteristics of students. GATES includes nine items about general artistic ability. Unfortunately, classroom teachers' knowledge of their students' musical, dramatic, or general artistic behaviors or activities in or outside of school is often limited. Even music teachers may lack the training or adequate knowledge of the students to make such judgments. These processes may succeed as a means to raise teachers' awareness of and value for artistic abilities. They are not, however, intended to be reliable assessments on their own.

Appropriate Assessments for Specific Purposes

The validity and equity problems in existing assessment and selection processes in the arts do not negate their effectiveness for the specific purposes for which they were designed. Administered by teachers trained in their use, the processes may efficiently fulfill their function of selecting, screening, or identifying specific artistic behaviors and characteristics. A brief screening process is often all that is feasible in order to offer opportunities for as many students as possible to participate in a special program or performance. A style-specific assessment may be valid if all of the students have an equal background in that style. These processes should not, however, be considered complete or accurate assessments of potential talent. When talent is defined and assessed too narrowly, many students will be missed, many more discouraged, and the conception of artistic talent will remain isolated from other abilities and intelligences.

Design Elements of Valid Performance Assessment

Authentic Arts Experiences. To overcome the kinds of problems listed above and provide a sound system of assessment for a wide variety of settings and student populations, the process must provide authentic arts experiences in an environment that allows artistic behaviors to emerge. The authenticity and artistic integrity of the activities are ultimately the primary challenge to the validity of any arts assessment. If the artistic experience is not authentic, then the students' responses are unlikely to be artistic. Authenticity of an arts experience encompasses the characteristics summarized by Wiggins (1998) and includes additional aspects particular to the arts. Authentic assessment tasks, according to Wiggins, reflect the context in which adults perform and are tested. The tasks must be active, requiring students to apply skills and knowledge to work on complex, realistic, real-world tasks in the discipline, while allowing opportunities for rehearsal, practice, and feedback (Linn & Baker, 1996). In the performing arts, tasks are most often performed in groups, requiring students to work together, watch, listen, and respond to each other. The arts require time for physical and mental preparation and warm-up and an atmosphere in which students can feel comfortable to take risks, communicate their feelings and ideas, and commit themselves fully to the activities (Eisner, 1994; Gardner, 1973). Most important, artistic tasks are open ended. According to Eisner, artistic tasks are a "process in which skills are employed to discover ends through action," as opposed to a craft, in

which skills are used to "arrive at preconceived ends" (p. 155). Time limitations, the potential stress and pressure of the testing environment, and the single-right-answer mentality are all severe threats to the authenticity of arts assessment.

Flexible Grouping. Any group-administered assessment faces particular reliability problems. Random assignment of subjects is impractical in schools, and the effect of intraclass correlations may be substantial in creating artificial findings (Huck & Cormier, 1996). However, the importance of establishing a safe, supportive environment among peers makes the classroom the most natural grouping (and, by far, the easiest to schedule). While not all classes are equally supportive, it is the peer group in which students tend to be most familiar and at ease. The use of intact classrooms also facilitates the involvement of the classroom teacher in the assessment process, which has many potential benefits.

Careful grouping of students within the arts assessment class is also important to the validity of the results. For assessors to provide complete and accurate scores for multiple students while observing them in complex, authentic arts activities, they need to be able to see and hear every student within various group sizes and with different partners. The physical arrangement of students and the ease and flexibility of changing groupings can have significant effects on both the ability of students in the class to perform at their best and on the assessors' ability to make perceptive judgments.

Skillful Facilitation. Effective assessment facilitation involves clear and simple verbal instructions and modeling; a positive, encouraging attitude; and the ability to observe the students while simultaneously directing activities. The facilitator must also know how and when to give appropriate feedback to students without biasing the assessment. Observing students' responses to suggestions and criticism is an important feature of authentic performance-based assessment (Wiggins, 1998).

Easy Scoring. The active, transitory nature of group arts activities makes scoring difficult. The scoring system and rubric must be simple to use, facilitating quick judgments. If the mechanics of scoring involves frequent interruptions, adds stress to the environment, or requires grouping students in ways that are unnatural or inappropriate to the activity, the authenticity of the activities will be compromised. Assessors need to be able to score students easily and continuously while keeping focused on the students.

In summary, previous arts assessment models are fraught with psychometric, practical, and cultural problems. Capable assessments, such as the models we now describe, contain a number of important features that enhance the dependability and fairness of the process.

Methods

D/M/T TAP has undergone three phases of testing. The initial tests of the process were conducted in three New York City Public Schools (1991–93), in which one art form was assessed in each school. After 3 years, the process was expanded to seven additional New York City schools (1994–96). The final stage involved two schools in Ohio in which all three art forms were administered to the same students (2001–03).

Sample. Schools were selected for the initial study based on their existing participation with ArtsConnection's Young Talent Program (YTP). YTP works in four of the five boroughs of New York City in schools with limited arts instruction, a high percentage of students receiving free and reduced lunch, and significant bilingual and special education populations. The three schools involved in the initial testing of D/M/T TAP were selected based on additional criteria of the Jacob Javits grant (inclusion in a HUD empowerment zone) and the willingness of the faculty to participate in ongoing professional development workshops over the course of the grant. The two schools in the Ohio expansion study were identified using the same general criteria of the Javits program and also represented broad geographic and ethnic diversity.

The assessments were conducted with the fourth grades in three schools in the initial study ($n = 639$) and with grades two through six in the expansion study ($n = 767$). All of the classes in a grade participated in the process. If every class could not be scheduled separately due to time limitations, small classes, such as self-contained special education or bilingual classes, were combined with each other or with other classes. Table 1 presents the demographics and academic test-score data for the identified students in the 10 YTP schools for the school year 1995–96.

Research Questions

The initial research focused on four major areas: content validation of the talent definitions and behavioral descriptors, discriminant

Table 1

Demographics and Test-Score Profile of Identified Students for 10 YTP Schools

Totals	*N*	Ethnicity					Gender		Test scores in quartiles (NCE)							
		His	Blk	Oth	Wht	As	F	M	1–25 Math/Read		26–50 Math/Read		51–75 Math/Read		76–99 Math/Read	
Identified students	389	119	212	33	21	4	226	163	28	51	82	72	81	75	35	36
%	100	31	54	9	5	1	58	42	17	22	35	33	33	29	15	16
Total Schools* %	1948	40	48	4	5	3	52	48	18	25	42	35	30	32	10	8

*Average of 10 schools based on average program year (1995–96)

validity comparing the results to other academic and affective measures and indicators, interrater reliability among the assessment panel (two artists and the classroom teacher), and the effectiveness of the process in selecting students who would be successful over time in a rigorous instructional program. The specific research questions were as follows:

1. Do the talent criteria developed for the process constitute a complete and coherent definition of talent in each art form?
2. Is the process equitable and independent of other measured variables? (i.e., Do students identified through the process reflect the school population in terms of gender and ethnicity? To what extent are scores from D/M/T TAP correlated with other measures of school performance and affective variables?)
3. Are the results of the process reliable? (i.e., Do the observers agree with each other and are their observations corroborated by independent experts? How many sessions are needed to arrive at a consensus between raters? Are classroom teachers able to assess their students reliably?)
4. Are students selected through the process successful in an advanced instructional program?

Description of D/M/T TAP

Overview of the Process. D/M/T TAP (Appendix A) was designed to assess systematically the artistic talents of all students, to identify students who are ready for advanced instruction in the art form, and to provide empirical data for designation of students as gifted and talented. A larger goal of the process is to raise appreciation and understanding of the artistic abilities and potentials of all students on the part of teachers, parents, and the students themselves. The current version of the assessment process was based on earlier assessment systems used in ArtsConnection's Young Talent Program between 1978 and 1991 (Briggs, 1991; BrooksSchmitz, 1990), as well as existing performance-based processes used in gifted education (Maker, 1992; Renzulli, Reis, & Smith, 1981; Torrance, 1966).

D/M/T TAP is administered to intact classes by a team of two trained arts instructors over a five-class series. Three assessors—the two arts instructors and a classroom teacher or specialist—rate all of the students on a written checklist of 8 music, 10 dance, or 4 theater items. Scoring is done on a simple notice/not notice scale for each

item, and total item scores are based on the sum of notices from all assessors. Each assessor also gives an overall, holistic score (1–5) for each student at the end of every class. After four classes, the item and overall scores from the three assessors are combined and averaged over the number of sessions attended. Students must attend a minimum of two sessions to receive a total score. Students are invited to a fifth "callback" session, based either on predetermined cutoff scores or by the number of students that can be accommodated in the callback session.

Definition of Talent. To create an assessment appropriate for use with diverse populations and flexible enough to be administered by a variety of arts teachers, ArtsConnection convened panels of arts educators in each art form representing a wide range of artistic styles and techniques (ArtsConnection, 1993, 1996). Drawing on their own professional and teaching experiences, the arts educators defined behaviors indicative of outstanding ability and the potential for development through instruction. The items and behavioral indicators defined by the panels transcend any single style and are articulated in simple, nontechnical language. Talent definitions are included in Appendix A.

The items in each art form can be grouped within three major categories that correspond to the Renzulli three-ring conception of giftedness (Renzulli, 1978). In Renzulli's conception, talent is an interaction of three clusters of traits: above-average ability, creativity, and task commitment. Ability must be above-average, but not necessarily prodigious. The student needs a basic proclivity toward an area to excel in it; but equal emphasis is put on the child's creativity and task commitment, which can lead to creative/productive accomplishment and the realization of gifted potential (Renzulli & Reis, 1997). Talent involves a set of behaviors that demonstrate the potential for high-level learning and achievement in a domain (Renzulli, 1978). The panels of artists independently verified the relevance of this three-part definition in their professional and educational experience.

Instructors and Observers. Two trained arts instructors administer the assessment, alternating between observing and facilitating so that at least one of the artists can record observations at all times. The use of an instructional team offers the potential for diversity in artistic and teaching style and a variety of viewpoints. Whenever possible, the instructional team consists of a male and a female, representing different cultures and artistic styles. The third assessor is

the classroom teacher or school specialist, who observes the entire class and has no teaching responsibility. Additional school personnel may act as observers, but only those who have been present for all sessions contribute scores to the assessment.

Assessment Activities and Curriculum Design. The curriculum for the five assessment sessions is designed by the assessment facilitators using the framework developed for the process. The framework defines the structure, groupings, and balance of various types of activities over the course of the process. The specific activities presented by the facilitators come from their own teaching practice and are designed to allow each of the behavioral indicators to be assessed in every session. Regardless of the particular artistic style or technique employed, the classes are designed to reflect authentic practices in the art form and engage students in complex activities requiring learning, problem solving, improvisation, cooperation, and the ability to take and use direction and feedback.

In dance, teams of artists from modern, West African, jazz, Caribbean, ballet, and creative movement traditions developed and facilitated assessment processes during the study. Music focused primarily on percussion and voice with artists from jazz, Orff, African, folk, and classical backgrounds. Theater instruction tends to be more uniform because most drama teachers share a background in improvisation, which, to minimize language and training differences, is a natural style to use in the assessment process.

The ideal situation of teaming two artists with different styles and cultural backgrounds is not always feasible. If the artists come from similar artistic backgrounds, it is essential that they vary the activities and the stimuli, including types of music, rhythms, or settings. A modern dancer need not try to teach African dance, for example, but will include some activities that emphasize rhythm, accompanied by strong percussion. Likewise, an African dance teacher will include some activities that allow students to use free or open rhythms and vary the musical accompaniment for the exercises. Both modern and African dance teachers will structure the class to allow approximately the same amount of time for warm-up, learning set movements, and improvisational activities and will vary the instructional mode (i.e., observation and following, verbal instruction, etc.). In music, the differences between vocal and instrumental music skills and the range of potential repertoire require the artists to provide activities that balance rhythm and melody, voice and percussion, familiar and unfamiliar songs, and creative improvisation. The biggest challenge in theater is to engage

students in activities that emphasize both verbal and nonverbal skills and that allow students to demonstrate collaborative and imaginative abilities that are not always as immediately apparent as individual speaking and acting skills. A diverse range of abilities involved in acting, playwriting, and directing can be seen in improvisation activities; but the process must be designed to allow subtlety, listening skills, and imagination to emerge and be assessed.

Training of Arts Facilitators and Teacher Assessors. The assessment facilitators can be school arts specialists, professional teaching artists, or other qualified arts teachers. Facilitators participate in a 4-day training process in which they familiarize themselves with the criteria and assessment framework, develop their own five-session assessment curriculum, and field test some of the activities with students. At the end of the training and a successful pilot administration, facilitators are certified to administer TAP in a particular art form (Oreck, 2002). Teacher assessors participate in a preassessment workshop to learn the criteria and scoring system and to help them prepare their students for the process. Immediately following each assessment session, the three assessors hold a 10-minute discussion while the students work quietly or are escorted to the classroom or library. Each student in the class is mentioned in the discussions. This serves to increase the assessors' awareness of all the students in the class and is an important part of the ongoing training of teachers.

Instruments and Psychometric Evaluation. Each assessor completes a checklist of observed behaviors (Observation Tally Sheet) for each session. When an observer notices one of the listed behaviors, a plus mark (+) is placed next to the relevant item in the student's box on the Observation Tally Sheet. In dance and music, one plus mark per rater for each item is counted toward the student's final item score for a session. In theater, which has only four items, two marks per rater per item are counted. The maximum score is the number of items marked by each of the three raters. Marks are not to be erased and negative marks are not scored. Additionally, each observer provides an overall holistic rating (1–5) for each session. This overall rating is combined with the item score for a final four-session total. Final scores are standardized by classroom and grade to rank students and to establish cutoff scores for inclusion in the fifth call-back session. Typical cutoff scores for official gifted and talented designation are $Z = +2.0$ and for selection for advanced instruction $Z = +1.0$. These levels can be adjusted for local gifted and talented program guidelines or to fill the available spaces in the instructional program.

Talent Development Opportunities. In the Young Talent Program, identified students participated in two classes per week taught by professional teaching artists from ArtsConnection's artist roster. Identified students also participated in 5–10 master classes at professional studios throughout New York City during the school day. Students were formally evaluated twice each year by the arts instructors. Parents or primary caregivers were provided with information and assistance to take advantage of local cultural institutions, scholarship opportunities, magnet schools, and summer programs. In Ohio's Project START ID, two classes per week, one during and one after school, were taught by school art and music specialists and professional teaching artists from community arts organizations.

Results

Research Question One: Definitions of Talent

Content Validity Evidence. Content validity evidence for the talent criteria was obtained during the development phase with a panel of experienced arts educators representing a variety of styles and teaching experience in each of the three disciplines. The final list of items and their behavior descriptors were then reviewed by other professionals in the art form, as well as by experts in gifted education and psychometrics.

Construct Validity Evidence: Factor Analysis. To study the dimensionality of the talent definitions, exploratory principal axis factor analyses were run for the observational ratings (items summed across the four sessions) in music, dance, and theater. In music, a single factor emerged (eigenvalue > 1.0), explaining 91% of the item covariation with a minimum loading of .75 (alpha estimate = .80). In dance, a single factor accounted for 89% of the covariation, with a minimum loading of .61 (alpha estimate = .78); theater ratings delivered a single factor that explained 97% of the ratings' covariation (alpha estimate = .86). Table 2 shows the loadings of each item with its factor.

Research Question Two: Discriminant and Convergent Evidence

Discriminant Evidence. Direct discriminant function analyses (DFA) were used to estimate the ability of D/M/T TAP to recognize characteristics of talent not measured by other tests and to ascertain

Table 2

Factor Loading for Dance, Music, and Theater Items

Dance		Music		Theater	
Item	Loading	Item	Loading	Item	Loading
Coordination	.90	Perception of sound	.91	Imagination	.89
Memory and recall	.89	Ability to focus	.90	Collaboration	.78
Physical control	.88	Expressiveness	.89	Physical	.76
Movement qualities	.86	Rhythm	.88	Focus	.68
Ability to focus	.82	Enthusiasm	.82		
Expressiveness	.82	Coordination	.81		
Rhythm	.81	Composition/improvisation	.79		
Spatial awareness	.80	Perseverance	.75		
Improvisation	.72				
Perseverance	.61				

the extent to which the identified students reflect the demographic makeup of the school population. It was hypothesized that the selection of students should be based on their performance during the assessment classes, rather than other factors, such as academic performance, measures of self-concept, ethnic group, or gender, factors, frequently correlated with participation in gifted and talented programs. The predictor variables were gender; ethnicity (dummy coded); and the Metropolitan Achievement Tests Math NCE (normal curve equivalent) scores (MAT-Math; Prescott, Balow, Hogan, & Farr, 1985/86; Hogan, 1986); Degrees of Reading Power NCE scores (DRP; New York State Department of Education and Touchstone Applied Science Associates, 1981), and the Piers-Harris Self-Concept measure (Piers, 1984), which was administered to all participating students.

DFAs were performed to predict student status: Identified (n = 112), Waitlisted (n = 157), or Not Identified (n = 370). Only the TAP results were significant ($p < .001$) in predicting group membership. In dance, they explained 65% of the variation in group membership; in music, 61%; and, in theater, 55%. According to Cohen (1992), these are very large effect sizes. None of the other variables—academic test scores, self-esteem subtest scores, gender, or ethnicity—were significant predictors of eventual selection. Correlations between arts talent and reading and math scores ranged from .08 to .25 (r^2 from .01 to .06), showing little correlation between these constructs.

Reading scores for identified students ranged from the 2nd to the 99th percentile, with more than half (62%) falling into the bottom two quartiles, which approximated the test performance of the schools as a whole. Fewer than 10% of students identified through the TAP would have been recognized for gifted and talented programs based solely on academic test-score criteria. As shown in Table 1, the identified group was generally representative of the overall school demographics in terms of gender, ethnicity, and academic test scores and included students from self-contained special education and bilingual classes.

Convergent Evidence. All participating classroom teachers completed a preassessment predictive instrument prior to the first TAP session (Teacher Searchlist; Baum, 1990). The Teacher Searchlist asked teachers to identify students who possess gifted and talented potential in dance, music, theater, or other general areas of gifted behavior (e.g., learns easily, is curious and creative, is an avid reader, has deep interests, has spatial talents, shows leadership potential).

Table 3

Interrater Reliability Results for Initial Tests (Fourth-Session Overall Rating)

	Music (n = 227)			Dance (n = 192)			Theater (n = 134)		
Assessor	A	B	C	A	B	C	A	B	C
A-Artist	--			--			--		
B-Artist	.654	--		.782	--		.716	--	
C-Teacher	.672	.788	--	.813	.817	--	.737	.545	--

Table 4

Interrater Reliability Results for Expansion Study (Fourth-Session Overall Rating)

	Music (n = 345)			Dance (n = 337)			Theater (n = 356)		
Assessor	A	B	C	A	B	C	A	B	C
A-Artist	--			--			--		
B-Artist	.859	--		.830	--		.718	--	
C-Teacher	.769	.851	--	.652	.707	--	.652	.736	--

As expected, the teacher prediction of talent in the arts was significantly correlated with eventual identification through TAP. For music, the average correlation was .40 (r^2 = .16, representing the overlap in variance between TAP scores and Searchlist ratings). In dance, the average correlation was .49 (r^2 = .24) and in theater, .42 (r^2 = .17). While statistically significant, less than one fourth of eventually identified students were recognized by their teachers as potentially talented prior to the assessment process.

Research Question Three: Interrater Reliability

In the initial research, mean interrater reliability estimates (intraclass correlation) among the three assessors across the audition process ranged from .55 to .74. Interrater reliability coefficients between artists and among the artists and teachers improved each session, reaching a moderately high level by session four (.67 for

Table 5

Average Fourth-Session Alpha Reliability Estimates (Overall Rating)

	Music	Dance	Theater
Initial study	.71	.80	.65
Expansion study	.80	.84	.87

Table 6

Stability Results for Expansion Study

	Music	Dance	Theater
Session 1–2	.539	.663	.480
Session 2–3	.579	.653	.352
Session 3–4	.561	.688	.550

music, .82 for dance, and .74 for theater). Table 3 summarizes the interrater correlations for the fourth session of the process in the initial study.

Interrater reliability estimates (intraclass correlation analysis) improved in all art forms in the expansion study, shown in Table 4. Alpha reliability estimates also improved in each subsequent administration of the process by the same group of assessors and between the original and expansion studies. Table 5 summarizes the alpha reliability estimates from the final administration of each study.

Stability. Stability estimates for the process were calculated over three separate intervals, and session-to-session estimates ranged from .35 to .68. Table 6 summarizes the stability estimates from one school over 1-week intervals.

Number of Sessions Needed. Seven assessment sessions were conducted in the initial tests. Interrater reliability reached a peak by the fourth session, and more than 98% of students who were eventually selected for advanced instruction had been identified by that point. Based on these results, the process for the expansion study was shortened to four sessions, with a fifth-session callback. Reliability estimates for the expansion study supported the four-session structure. Interim rankings after two and three sessions suggest that fur-

ther shortening of the process would result in as much as 30% misranking of students.

Research Question Four: Construct Validity Evidence

Contrasting Groups Evidence. Construct validity evidence was gathered by collecting new assessment ratings on students a year after the original assessment process. A random sample of Identified (n = 45) and Not Identified (n = 44) students participated in a new talent assessment conducted and rated by professional artists unfamiliar with the students, using the original talent criteria and activities that none of the students had previously done. A Hotelling T^2 was used to compare Identified and Not Identified students on all ratings simultaneously. Univariate t tests were used as a post hoc probe of the significant T^2. To protect against inflated Type I error rate, a Bonferroni correction was applied to the alpha value: The nominal alpha of .05 was divided by 8 (consecutive t tests) to give a per-comparison alpha of .006. The Hotelling t tests (df = 46) showed that Identified students dependably received higher talent ratings (T^2 = 36.88, p = .004). The univariate t tests favored Identified students for each of the rated behaviors.

Construct Validity Evidence: Informal Supplementary Data. During 2 years of advanced instruction, 82% of Identified students made good to excellent progress on written semiannual evaluations by arts instructors using the original identification criteria. The overall program attrition rate was below 8% for reasons other than leaving the school. Additional evidence of students' readiness included high attendance rates for after-school classes, the amount of home practice reported, and instructor's reports of students' on-task behavior during arts classes. Approximately one half (38 of 80) of graduating fifth and sixth graders in 1995–96 participated in some form of ongoing arts training, and graduates received scholarships to the Walt Disney Youth Orchestra (4), the Julliard School Saturday program (3), the Martha Graham School (2), Alvin Ailey American Dance Center (3), Dance Theater of Harlem (2), Teatro Del Opera (4), and Ballet Hispanico (5), among others (Oreck, Baum, & McCartney, 2000).

Discussion

These studies provide evidence for the validity and reliability of the TAP results in dance, music, and theater. The results of the

Ohio study suggest that the process is adaptable to different school settings and that the training procedures for artists and classroom teachers are adequate to replicate the process with similar results (Pearson, 2003). The students identified through D/M/T TAP, unlike those selected for gifted and talented programs through IQ or academic test scores, accurately represented the demographics of the schools, including students in self-contained special education and bilingual classrooms. The label *gifted and talented* is a hot-button topic for parents, administrators, and teachers throughout education. D/M/T TAP uses "current readiness for advanced instruction" as a functional definition of talent to attempt to defuse the issue and to put the focus of the assessment on serving, not simply labeling, students. It is challenging, to be sure, to follow students over time and track their progress in an ongoing advanced instructional program, but that is ultimately the best practical measure of the validity of the assessment.

Definition of Talent

The D/M/T TAP definitions of talent are based on the idea that artistic talent in each individual is a combination of many factors, not a single dominant skill or characteristic, such as the ability to match pitches, balance on one leg, or speak loudly. If artistic talent is manifested in the constellation of skills, motivation, and creative expression, then all of the items should be at least moderately correlated. The factor analysis results, in fact, revealed a single, relatively homogeneous factor in each art form. The observational ratings formed unidimensional composites with high loadings, demonstrating simple structure and little unsystematic variation. High correlations among items support valid and unified definitions of talent. It could be argued that, with such high correlations, the assessment could be conducted with fewer items. However, the goal of the process—to create a specific and detailed talent profile for every student—implies using the most complete list of criteria that is feasible.

The diverse panels of professional artists and arts educators, who defined the behavioral definitions and the diverse instructional experiences in which selected students successfully participated, provide strong evidence to support the concept that talent transcends specific artistic styles, techniques, and cultural backgrounds. Given adequate experiences, students will eventually make a choice of styles or instruments on which to focus. Thus, the wider the

range of artistic experiences to which students are exposed, the more likely it is that their potential talent can be realized.

Reliability and Stability Results

Just as no single criterion defines talent, the process relies on the collective observations of multiple assessors, rather than a single assessor. Given the complexity of the activities and the difficulty of observing large groups simultaneously, it is expected that assessors would notice different people and different behaviors at any given moment. The reliability of the results is thus based on the sum of observations and the strength and consistency of the student's performance in a range of activities over time, which reflect actual practices in arts classes.

Interrater reliability estimates increased in each subsequent administration of the process with the same teacher/artist teams, likely due to both improved observation skills and more effective facilitation on the part of the artists. The high level of agreement between classroom teachers and the arts facilitators is an important finding supporting one of the central purposes of D/M/T TAP: to raise teachers' appreciation of and support for the artistic abilities of their students. Before the first assessment process began, teachers were able to predict fewer than 25% of the students who were eventually identified for advanced instruction. Yet, although most of the teachers had little previous experience or prior expertise in the art form (ArtsConnection, 1996; Oreck, 2001), they were as able to recognize talented behaviors in their students as the arts experts did after just a few TAP observations. While some teachers may initially have been influenced by the artists (a training effect), artist-to-artist agreement was not significantly higher and, in some cases, was lower than artist-teacher agreement. Further studies revealed that the recognition of students' artistic strengths is a primary motivation for teachers to employ arts processes in their classroom instruction (Oreck, 2000).

Because each of the four sessions is designed to highlight different aspects of talent, it could be expected that different students would stand out in different sessions and that session-to-session stability estimates would, thus, be lower than interrater reliability estimates. Inconsistent performance over time and over various activities reveals valuable information about the students' readiness for advanced instruction. The relatively higher interrater reliability estimates suggest that the session-to-session differences are more likely the result of changes in student performance as opposed to inconsistency in assessor ratings.

Artistic Talent and Academic Achievement

Based on anecdotal evidence from teachers in the 12 years before the first Javits grant research, it was clear that a significant percentage of students identified as potentially talented had academic or behavior problems that put them at risk for school failure. The comment, "You've picked all of my worst kids!" was a frequent first response from teachers when they did not directly participate in the assessment. While there were many high-scoring students selected as well, it was the low academic achievers who most impressed their teachers. The current studies verified the anecdotal observations. Low correlations found between reading and math scores and D/M/T TAP results (r^2 = .03 - .06) supported the theoretical divergence between achievement on academic tests and artistic talent. The difficulties that many high-ability, high-energy creative students encounter in test taking is borne out by the large percentage (more than 60%) of identified students scoring below grade level on reading tests in both the New York and Ohio studies.

These results show that the inclusion of arts talent in gifted and talented assessment can be a means to increase ethnic and cultural diversity and to expand representation of students from all academic levels and classrooms (including bilingual and special education) in the school. There is a danger, however, of creating a separate, parallel program for artistically talented students that is considered "nonacademic." In order to avoid this result, the message needs to be reinforced to teachers, parents, and students that there are many ways to be smart and that students who excel in rigorous arts instruction have the ability to excel in school. The primary importance of the classroom teachers' participation in D/M/T TAP is to help them appreciate the creative abilities, multiple intelligences, effective learning behaviors, and preferred learning styles of students and to use that knowledge to improve overall school performance (Baum et al., 1996). Teachers not only have the opportunity to see their students in a different light during the assessment, they observe the teaching methods and activities of the artists that inspired and allowed some of those "worst kids" to shine.

Prediction of Future Performance

One of the most significant markers of a capable talent assessment is its ability to predict future performance in a rigorous instructional program. The low attrition rate in the in-school, after-school, and studio master classes in the Young Talent Program is especially impressive, given the high mobility of students in the program schools and the

number of students experiencing academic, behavioral, and attendance problems, as well as extraschool problems. While this research design could not fully investigate predictive validity through the use of randomized placement, alternative assessments, or alternative instructional experiences, the process proved effective in selecting students who would be successful in the instructional programs offered in the various program schools and professional studios. The attrition rate dropped significantly when using the new processes, compared to the one-time, artist-only assessment process used previously. Between 1980 and 1990, fewer than 30% of the students continued in the Young Talent Program beyond the second year of instruction (BrooksSchmitz, 1990). Using D/M/T TAP, more than 70% of the students completed 3 years of the instructional program.

Additional evidence gathered in the blind ratings of matched pairs of selected and not-selected (including wait-listed) students 2 years after the original assessment process supported the accuracy of the original selection. Not surprisingly, given the increased instruction offered the selected students (25 weeks of 45–115 minutes per week in school vs. 3–5 weeks for 45 minutes for not selected), the results showed large differences between selected and not-selected students on all criteria (dance, $t = 5.40$; music, $t = 5.32$; theater, $t = 4.28$). Each of these differences favored the selected students. Most strikingly, differences between selected and wait-listed students, very small after the initial process, had widened so that wait-listed were virtually indistinguishable from not-selected students by sixth grade. This shift reaffirms the need to identify and develop potential talents of all students, and it suggests further studies of artistic development with both identified and not-identified groups to investigate a "use it or lose it" hypothesis.

TAP Scoring System

The two-part scoring system was designed to balance specific and general observations of students in each assessment session. Specific traits and behaviors are scored with a simple checklist. In earlier studies, complex rubrics requiring assessors to make fine qualitative distinctions were found to be more cumbersome and less reliable than the simple summing across various activities we had developed. The overall 1–5 rating allows assessors to synthesize their cumulative professional judgment about students' readiness for advanced instruction. It also seeks to lessen potential ceiling effects of the scoring system that would tend to undervalue students who are exceptional in one or just a few categories. While the theoretical basis of

the talent definitions indicates that talent is the combination of many behaviors reflecting above-average ability, motivation, and creativity, the talent profile is different in each individual. There are students who receive few behavior notices, but whose unique creative improvisations or intense determination indicates considerable artistic potential. This is borne out in the experience of many professional artists who have compensated for apparent deficiencies in some specific abilities or skills with a high level of task commitment or creativity (Bloom, 1985; Subotnik, 1995). The expanded range of the combined item and overall scores can also partly overcome initial low scores for students who are at first shy or withdrawn, but improve dramatically as they become more comfortable in the activities. The item and overall scores, of course, tend to be highly correlated (mean *rs* = .78 - .93), but the overall scores tend to have higher interrater reliability and stability estimates.

Postclass Discussion

Some psychometricians may argue that the postclass discussion threatens the independence of the assessors' scores in subsequent sessions. The goal of the process, however, is to complete the most thorough assessment possible for every student in the class, and the discussion immediately alerts the assessors to students they have missed. The strongest and weakest students in the class tend to be easily and frequently noticed by everyone. It is those in the middle, students who are quiet or shy or who need a little extra time to understand instructions, who are most likely to be overlooked. Further, sharing this information helps assessors disclose their own biases and preferences by requiring them to articulate the specific behaviors they have noticed and to hear the observations of others. We consider incomplete or biased assessments to be stronger validity threats than the problem of assessor independence. Further, the discussion is an important component in training classroom teachers to understand the criteria, vocabulary, and approaches used by the artists; and it allows the teachers to share with the artists potentially relevant background information about students.

The Role of Culture and Artistic Styles in the Assessment Curriculum

Culture cannot be removed from the assessment process. Talent is a cultural-dependent construct, and art is embedded in a cultural context. The lack of predictive validity reported for existing arts assess-

ments results from the weak link between the characteristics being measured and authentic practices in the art form. Although student responses are greatly affected by the specific teacher and artistic style being taught, the most important aspect of the assessment is the complete engagement of the students in authentic arts activities. In order to reach that level of engagement, the facilitators must be comfortable with what they are doing, clear in their instructions and feedback, and motivating to the students. They are generally most able to achieve these goals in the style of their own expertise, rather than teaching a set of mandated exercises or decontextualized material. This approach is very much like participant observation in qualitative intervention research where the investigator assists in creating a tailored intervention, rather than handing over a preplanned structure that fits no one in particular. The curricular framework requires a range of types of activities, teaching modalities, groupings of students, and opportunities for students to improvise and solve open-ended problems. Further, while each session has a different central focus, each class is designed to assess all of the behavioral items. The focus on broad talent definitions and varied activities has resulted in a process that can be adaptable and responsive to local norms while maintaining fair and appropriate assessment procedures.

The experiences of YTP students in new and unfamiliar styles and techniques through the professional master class series and post-YTP arts training pursuits provides additional evidence for the generalizability of potential talent beyond specific techniques or culturally based styles (Oreck et al., 2000). In one example, the four students who were selected to attend the Disney Youth Orchestra had not played classical music and did not read musical notation before being selected for that weeklong experience. Their natural talent and training in Orff, jazz, and African percussion prepared them to pick up the material easily; and two were chosen to perform lead xylophone and timpani parts in the televised finale. Likewise, a number of students who had primarily studied African and Afro-Caribbean dance at their schools were awarded scholarships to the Dance Theater of Harlem (ballet), the Martha Graham school (modern), Ballet Hispanico (flamenco and ballet), and the Alvin Ailey American Dance Center (modern and ballet).

Conclusions

In a time of decreasing budgets and intense focus on testing and basic proficiency, both the arts and gifted education are being

pushed ever further to the educational margins. For arts talent assessment processes to be more widely adopted in schools, research must show both the positive impact and empirical validity of such processes on students and teachers.

The difficulty of defining and assessing artistic potential and developing evidence-based assessment processes has contributed to the continued undervaluing of the arts in schools. If administrators, teachers, and parents can gain reliable information on students' artistic potentials and the particular skills and learning behaviors developed through the arts, they are more likely to see the benefit of basic arts instruction for all students. Armed with supporting data, schools and parents may also become more motivated to provide advanced instruction for those students who are ready; pursue links with local community arts institutions; include more arts experiences in gifted programs; and use performance-based assessment data to help talented, low-scoring students improve their academic performance. Given the current financial realities in schools and the continuing cutbacks in arts programs, arts assessments must be linked to overall school improvement, or they are unlikely to be employed. Our research has shown that, beyond the goal of identifying outstandingly talented students, careful and systematic arts assessment can provide valuable information about the strengths, interests, and learning styles of every child.

References

Abeel, L., Callahan, C., & Hunsaker, S. (1994). *The use of published instruments in the identification of gifted students.* Washington, DC: National Association for Gifted Children.

Arnheim, D., & Sinclair, W. (1974). *Basic motor ability tests*. Long Beach: California State University, Institute of Sensory Motor Development.

ArtsConnection. (1993). *Talent beyond words* (Report to the Jacob K. Javits Gifted and Talented Students Education Program of the U.S. Department of Education, Office of Education Research and Improvement, #R206A00148). New York: Author.

ArtsConnection. (1996). *New horizons* (Report to the Jacob K. Javits Gifted and Talented Students Education Program of the U.S. Department of Education, Office of Education Research and Improvement, #R206A30046). New York: Author.

Baum, S. M. (1990). Recognizing talent in young children. In S. Baum, S. Reis, & L. Maxfield (Eds.), *Nurturing the gifts and tal-*

ents of primary grade students (pp. 30–40). Mansfield Center, CT: Creative Learning Press.

Baum, S. M., Owen, S. V., & Oreck, B. A. (1996). Talent beyond words: Identification of potential talent in dance and music in elementary students. *Gifted Child Quarterly, 40*, 93–101.

Baum, S. M., Owen, S. V., & Oreck, B. A. (1997). Transferring individual self-regulation processes from arts to academics. *Arts Education Policy Review, 98*, 32–39.

Bloom, B. S. (Ed.). (1985). *Developing talent in young people.* New York: Ballantine Books.

Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences.* New York: Schirmer.

Briggs, D. P. (1991). *Young talent: The first decade.* Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

BrooksSchmitz, N. (1990). *Young talent research project: An analysis of the effect of arts-in-education programming on the motivation, academic performance, and personal development of inner-city youth involved in the young talent program*. Unpublished manuscript, Teachers College, Columbia University, New York.

Byrnes, P., & Parke, B. (1982, April). *Creative Products Scale: Detroit public schools*. Paper presented at the annual meeting of the Council for Exceptional Children, Baltimore.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Darling-Hammond, L. (1994). Performance-based assessments and educational equity. *Harvard Educational Review, 66*, 5–30.

Eisner, E. W. (1994). *Cognition and curriculum reconsidered* (2nd ed.). New York: Teachers College Press.

Elam, A., & Doughty, R. (1988). *Guidelines for the identification of artistically gifted and talented students* (Rev. ed.). Columbia: South Carolina State Department of Education.

Gardner, H. (1973). *The arts and human development.* New York: Wiley & Sons.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: BasicBooks.

Gilliam, J., Carpenter, B., & Christensen, J. (1996). *Gifted and talented evaluation scales.* Austin, TX: PRO-ED.

Gordon, E. E. (1965). *Musical aptitude profile.* Boston: Houghton Mifflin.

Gordon, E. E. (1979). *Primary measures of music audiation and intermediate measures of music audiation.* Chicago: GIA.

Haroutounian, J. (2002). *Kindling the spark: Recognizing and developing musical talent.* New York: Oxford.

Huck, S. W., & Cormier, W. H. (1996). *Reading statistic and research* (2nd ed.). New York: HarperCollins.

Kay, S. I., & Subotnik, R. F. (1994). Talent beyond words: Unveiling spatial, expressive, kinesthetic, and musical talent in young children. *Gifted Child Quarterly, 38*, 70–74.

Linn, R. L., & Baker, E. L. (1996). In J. B. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities* (Yearbook of the National Society of Education, Part 1, pp. 84–103). Chicago: University of Chicago Press.

Love, J., & Kipple, B. (1995). *Arts participation and race/ethnicity: An analysis of 1982, 1985 and 1992 SPPA surveys.* Washington, DC: National Endowment for the Arts.

Maker, J. C. (1992). Intelligence and creativity in multiple intelligences: Identification and development. *Educating Able Learners, 11,* 12–19.

Marland, S. P., Jr. (1972). *Education of the gifted and talented: Report to the Congress of the United States by the Commissioner of Education.* Washington, DC: U.S. Office of Education.

New York State Department of Education & Touchstone Applied Science Associates. (1981). *Degrees of reading power.* New York: College Board.

No Child Left Behind Act of 2001, Pub L. No. 107–110.

Ohio Department of Education. (2000). *Project START ID: Statewide arts talent identification and development project.* Columbus: Author.

Oreck, B. (2000, April). *Artistic CHOICES: How and why teachers use the arts in the classroom*. Paper presented at the annual meeting of the American Education Research Association, New Orleans.

Oreck, B. A. (2001). *The arts in teaching: An investigation of factors influencing teachers' use of the arts in the classroom.* Unpublished doctoral dissertation, University of Connecticut, Storrs. (ProQuest cat. # 9999695)

Oreck, B. A. (2002). *D/M/T TAP training manual.* New York: ArtsConnection.

Oreck, B. Baum, S., & McCartney, H. (2000). *Artistic talent development for urban youth: The promise and the challenge*. Storrs: University of Connecticut, National Research Center for the Gifted and Talented.

Pearson, L. (2003). Identifying and serving our gifted students: How professional development can help meet the challenges. *TRIAD: Journal of the Ohio Music Education Association,70*(6), 37–41.

Piers, E. V. (1984). *Piers-Harris Children's Self-Concept Scale, revised manual.* Los Angeles: Western Psychological Services.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1985/86). *Metropolitan achievement tests* (6th ed.). San Antonio, TX: Psychological Corporation.

Profile: Eliot Feld. (1998, April, 16). *Parade Magazine*, p. 22.

Rainbow, E. L. (1965). A pilot study to investigate the constructs of musical aptitude. *Journal of Research in Music Education, 13*, 3–14.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*, 180–184, 261.

Renzulli, J. S., & Reis, S. M. (1997). *The schoolwide enrichment model* (2nd ed.). Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., Reis, S. M., & Smith, L. H. (1981). *The revolving door identification model.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., & Hartman, R. K. (1976). *Scales for rating the behavioral characteristics of superior students*. Mansfield Center, CT: Creative Learning Press.

Richert, S. E. (1992). *Equitable identification of students with gifted potential.* (ERIC Document Reproduction Service No. ED366159)

Roach, E., & Kephart, M. (1966). *Purdue perceptual motor survey.* Columbus, OH: Merrill.

Saunders, R. J., & Schmidt, L. (Eds.). (1979, March). *Con-Cept IV: Task force report on identifying the talented in the creative arts*. Hartford: Connecticut State Department of Education.

Seashore, H. (1938). *Psychology of music*. New York: McGraw-Hill.

Shwedel, A. M., & Stoneburner, R. (1983). Identification. In M. B. Karnes (Ed.), *The underserved: Our young gifted children* (pp. 17–39). Reston, VA: Council for Exceptional Children.

Sternberg, R. (1988). *The triarchic mind*. New York: Penguin Books.

Sternberg, R. J., & Grigorenko, E. L. (2002). The theory of successful intelligence as a basis for gifted education. *Gifted Child Quarterly, 46*, 265–277.

Subotnik, R. (1995). Talent developed: Conversations with masters in the arts and sciences. *Journal for the Education of the Gifted, 18*, 440–466.

Torrance, E. (1966). *Torrance tests of creative thinking.* New York: Dodd, Mead & Co.

U.S. Department of Education. (1995). *Arts education in public elementary and secondary schools.* Washington, DC: National

Center for Education Statistics. (Office of Educational Research and Improvement No. NCES 95-082)
Webster, P. (1994). *Measure of creative thinking in music*. Evanston, IL: Northwestern University Press.
Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

Appendix A
Dance Talent Items and Behavioral Descriptors

Skills	Motivation	Creativity
1. Physical control knows by feeling can make adjustments can balance on one leg has strength in legs, arms, torso can maintain corrections **2. Coordination and agility** can combine movements executes complex locomotor patterns can isolate body parts from each other moves freely through space moves quickly **3. Spatial awareness** is aware of other people adjusts to other dancers and the space evens up the circle or line is accurate in time and space **4. Observation and recall** remembers information can perform without following can see and replicate movements accurately can build sequences **5. Rhythm** puts the beat in the body repeats rhythmic patterns accurately anticipates, waits for proper moment to begin can find the underlying pulse or beat	**6. Ability to Focus** directs attention makes full commitment to the movement is interested and involved in class **7. Perseverance** doesn't give up easily practices improves over time takes time to think tries hard to get it right	**8. Expressiveness** shows pleasure in movement performs with energy and intensity is fully involved communicates feelings **9. Movement qualities** displays a range of dynamics has facility moving in levels, directions, styles communicates subtlety moves fully connects body parts **10. Improvisation** responds spontaneously uses focus to create reality shows the details gives surprising or unusual answers

Music Talent Items and Behavioral Descriptors

Skills	Motivation	Creativity
1. Rhythm puts the beat in the body is able to sustain an even beat replicates rhythmic patterns accurately can play repeating patterns anticipates, waits for proper moment to begin can find the underlying pulse or beat **2. Perception of Sound** perceives differences in tone and pitch responds to dynamics can match pitches can replicate melodic phrases is able to sustain an independent part **3. Coordination** moves easily through space able to do two or more things at the same time can control body in movement and freeze sustains repeating patterns works with both hands	**4. Enthusiasm** responds joyfully eager to participate curious, asks questions is open to unfamiliar styles of music **5. Ability to Focus** directs attention makes full commitment to the task is interested and involved in class activities listens actively and carefully follows instructions **6. Perseverance** doesn't give up easily improves over time takes time to think is able to take and use corrections	**7. Expressiveness** responds with sensitivity performs with energy and intensity is fully involved communicates feelings **8. Composition and Improvisation** improvises spontaneously takes risks makes surprising or unusual statements creates sounds in original ways makes up songs

Theater Talent Items and Behavioral Descriptors

1. **Physical Awareness**
   responds with whole body
   is in control of body parts
   uses and perceives vocal qualities
   can use voice flexibly
   wants to be heard and understood
   is aware of space
   notices details
   observes carefully
   seems relaxed
   is unembarrassed
2. **Focus/Commitment**
   gives energy
   takes risks
   participates fully
   perseveres
   focuses eyes on the imagined environment and other players
   recalls instructions
   can revise and improve own work
3. **Collaboration**
   works with others
   responds to the audience
   accepts the "rules" of the exercise
   listens to teachers and peers
   takes direction and criticism well
   gives helpful suggestions
   takes a leadership role
4. **Imagination**
   offers ideas
   comes up with original or unusual suggestions
   finds multiple solutions
   makes the situation "real"
   solves problems
   sees the whole picture
   invents dramatic situations
   has a sense of effective timing