

Planning for Quality 102

Standards and/or Goal Development and Assessment

10/27/17

Assessment Terms and Definitions

Prepared for *Ingenuity*

Douglas G Boughton
Kerry J Freedman
Northern Illinois University

Evaluation in education is generally regarded as the appraisal of the value of an educational artifact requiring judgments to be made about the quality of a performance, artifact, or product. In the United States, the term is typically used in a broad way to refer to judgments of quality in any educational situation where determinations of merit are required. In other countries, such as United Kingdom, the term is frequently used more specifically to refer to the judgments about the quality or effectiveness of curriculum (or program).

Assessment is a term often used in Europe, the UK, and Australasia to refer specifically to the determination of levels of student performance in national or centralized systems of education, usually as an end point judgment in high-stakes situations where specific agreed procedures are employed at the state or national level to quantify the quality of student performance. In the USA assessment and evaluation are frequently used interchangeably. (Boughton, Eisner, and Ligtvoet 1996).

Testing requires an individual to demonstrate knowledge by performing tasks in response to a set of questions or completion of a defined exercise. Tests, therefore, are methods for securing information about learning, but they have no intrinsic connection to evaluation. A test simply provides information about which judgments have to be made. Typically, the results of a test are measured or judged to produce a score or grade (Eisner 1996, p.75). Testing is not uncommon in the arts. History of the arts is typically tested using traditional essay, or short answer methods. Studio practice and performance, on the other hand, is almost always practical in nature. In some contexts, performance tests take place in a single setting, such as a drawing exam, musical performance, and in others the test may extend over many weeks.

Grading is the use of a letter or symbol to represent a level of achievement following the judgment of evidence including artifacts, performance, or test scores made by a teacher/examiner. Grading is a reductionist exercise and is not the same thing as evaluation (Eisner 1996, p.75). A student who receives a “B”, for example, will understand their work is not as good as someone who received an “A”, but they will not know why unless further assessment information is provided, usually in verbal form. The explanation of grade is often conducted in arts classes using critique methods.

Measurement is the process of quantifying information that can be used to consider in judgments about student learning. A measurement activity is designed to determine whether something is present or absent in an object or performance. For example, counting the number of exploratory sketches in a student’s portfolio or the number of

incorrect notes played in a recital. Measurement activities in the arts tend not to be particularly helpful with the determination of artistic quality. If a student produces ten required exploratory sketches for inclusion in his/her portfolio it is not necessarily the case that the work is qualitatively better than another student who produced only six sketches rather than the required ten. The quality of sketches is not necessarily related to the number present in the portfolio.

Quantitative Assessment is closely related to measurement in that it employs a process of assessing student achievement based on things that can be counted. For example, the number of right answers in a test, the number of artworks completed, or the number of correct technical process employed in a performance task. The assumption underpinning this kind of assessment is that “more is better”. In the arts, this assumption falsely connects quantities with quality and is, for the most part, unhelpful in determining the virtue of students’ creative production. That said, there are some conditions under which this kind of assessment data can contribute to an understanding of student progress (such as the demonstration of knowledge of facts about art history) but for the most part does not address the fundamental issues related to artistic/critical thinking and creative production.

Qualitative assessment, on the other hand, is the process of making a judgment about the degree to which qualities are present in a performance, or object relative to an established standard. For example, evocative form, technical skill, or imagination. Such judgements are complex and require experienced assessors with intimate knowledge of the media employed, artistic genre, and student development to be able to make these judgments effectively. In contrast to quantitative judgments which rest upon binary judgments (artifacts that are either present or not present), qualitative assessments require judgments about the degree to which desired qualities exist. Such judgments provide a far more demanding task for the teacher, but are considerably more appropriate to assessment of learning in the arts.

Diagnostic assessment may be conducted as part of the day-to-day routine of teaching for the purpose of identifying students’ strengths and weaknesses. This form of assessment can include observation, testing, and analysis of classroom work and the intention is to inform teachers about the profile of student performance and provide input to improve future teaching strategies. Such assessments are extremely important in the visual arts since prior art learning experience is very often diverse and unpredictable as the years of schooling increase.

Formative Assessment provides information on a continuous basis to facilitate students’ understanding of their progress towards achievement of overall program and personal goals. This form of assessment is idiographic in nature and intended to inform individual students how they are progressing in their work. Such information is personal and does not typically reference the performance of others. In the art and design class such assessment is crucial since the expectation is for each student to follow their own pathway to an original personal outcome. The nature of art teaching requires teachers to engage in one-to-one discussions during which critique and encouragement are provided. Formative assessment tends to be a private affair between the teacher and the student unless such advice is given in a class group critique.

Summative assessments, sometimes called *gatekeeping assessments*, are an end point judgment about the degree of achievement evident in student work at the end of a substantial unit of study, such as a term, semester, or year. These assessments determine access for students to educational opportunities such as the ability to proceed further in their course of studies, entry to college or university, or final graduation. These assessments are substantially different from formative assessments since students' work is compared against the learning criteria established for the program (criterion-reference), and/or against the level of performance of other students (norm reference). Often these assessments employ rigid rules governing judgment, comparative benchmarks, and moderation procedures designed to ensure equivalents of judgment among students in different school settings.

High-stakes tests are, in some respects, similar to gatekeeping assessments for students, but, since 2001 (in the USA) the term has also embraced the kind of tests administered at national or state levels to assist administrators to make decisions about important, sometimes life changing consequences for the test taker and/or the educational system that administers the test. At the system level schools, or even states, may be at risk of closure or penalty if a significant percentage of test takers do not meet system defined standards. In the United States, this kind of testing was implemented following the "No Child Left behind" initiative introduced by the Bush administration in 2001.

Authentic assessment determines student performance in terms of the capacity of learners to complete a real-life performance task. For example, a student may demonstrate his or her ability to make a realistic drawing of a toy from observation. If that student was asked to answer multiple-choice questions about how to draw the toy this would be an inauthentic assessment of their ability to successfully complete that task. That is, they may answer the questions correctly but still be unable to complete the drawing task with any level of success.

A criterion in the context of arts education is the expression of the quality sought when a judgment must be made about something. eg creativity: "The degree of creative thinking demonstrated in the art work". "Criteria" is the plural form of the word ie one criterion or many criteria. The process of assessment in visual arts requires the expression of explicit criteria to guide the judgment of those charged with the responsibility of determining whether or not student products or performance contain the qualities that demonstrate a high level of learning, or "excellence" of achievement, in the context of the curriculum. Many common criteria include imagination, technical skill, use of formal qualities to achieve expressive outcomes, and so on.

A rubric is a set of statements describing performance levels that may be achieved in relation to a criterion. For example, "technical skill" is a criterion whereas a statement such as "Demonstrates a high level of skill in the use of charcoal using delicately articulated value gradients to create the illusion of space and form" could be a statement that describes the highest level of performance against the criterion "technical skill" for a specific project (e.g. charcoal rendering). A rubric typically will be comprised of two parts, a numerical scale and matching statements of performance levels. In the visual arts a three to five-point scale is typically used since fine-grained distinctions of quality are very difficult to achieve reliability using more than five levels. A rubric is based upon the assumption that any qualities that can be seen to be present in an object or performance are present to some degree. The following example (Table 1) shows how a

simple three level rubric for a class project might be developed using the “technical skill” criterion referenced above.

Table 1
Analytic Rubric: Criterion Technical Skill

Performance Level	Performance Descriptor
3 (Highest Level)	Demonstrates a high level of skill in the use of charcoal using delicately articulated value gradients to create a realistic illusion of space and form .
2 (Medium Level)	Demonstrates a moderate level of skill in the use of charcoal using a restricted range of value gradients to create a limited illusion of space and form .
1 (Lowest Level)	Demonstrates a minimum level of skill in the use of charcoal using a very limited range of value gradients to create an unconvincing illusion of space and form .

Analytic assessment employs separate criteria (and related rubrics) which each provide a separate number which are aggregated to form a total score for the artwork being judged. For example, a work may be judged against the criteria of “technical skill”, “expressive use of visual qualities”, and “degree of imagination” expressed in the work. A score could be given on a 1 to 5 scale for each of those criteria. Those three numbers are then added together to arrive at a final score for the piece. The assumption in such a practice is, of course, that each of the criteria carry equal weight in the determination of the value of the piece. (NOTE The individual technical qualities sought are color coded above).

Holistic assessment strategies, on the other hand, demand a single overall judgment. The examiner will refer to all the criteria, determine the manner in which each are manifested in the work, consider whether the genre of the work demands additional consideration beyond those defined by the criteria, and then provide a single score to express overall quality. The benefit of the holistic method is that analytic criteria used to determine qualities present in any given work are seldom mutually exclusive. Take, for example, the three criteria mentioned above, “technical skill”, “expressive use of visual qualities”, and “imagination”. Each of these criteria are not distinct from each other. For example, technical skill may be used in an imaginative way to achieve expressive outcomes. Separate judgments for each is difficult, and more likely than not is also inappropriate. In addition, each criterion may not be equally important given the nature of the work under investigation. Aggregation of three scores based upon a five-point scale may not accurately reflect an appropriate judgment the overall value of the work. It is also important to note that there is no difference in reliability between analytic and holistic judgments provided the examiners are experienced. An example of a holistic rubric to be used for a summative assessment of an art student’s portfolio follows:

Assessment (Content) Validity is achieved when the measures used to assess an art work or performance interrogate the essence of the subject matter. If the measures reveal what is held to be essential to the subject then the instrument or method can be regarded as valid.

Table 2

**Holistic Rubric
(Adapted from the International Baccalaureate)**

Select the descriptor below which best reflects the candidate's exhibition of work.

LEVEL	HOLISTIC DESCRIPTORS
5	A very powerful collection of work has been produced which demonstrates an exceptional understanding of the conceptual and technical underpinnings of artistic expression representative of the cultural context and chosen artistic genre(s). The work overall illustrates a highly sophisticated exploration of ideas appropriate to the visual arts, and an outstanding resolution of concept, media, and technical expression.
4	A strong collection of work has been produced which demonstrates a very good understanding of the conceptual and technical underpinnings of artistic expression representative of the cultural context and chosen artistic genre(s). The work illustrates a comprehensive exploration of ideas appropriate to the visual arts, and a good resolution of concept, media, and technical expression in the majority of works.
3	A reasonable amount of work has been produced which demonstrates an adequate understanding of the conceptual and/or technical underpinnings of artistic expression representative of the cultural context and chosen artistic genre(s). The work illustrates a reasonably focused exploration of ideas appropriate to the visual arts, and a good resolution of concept, media, and technical expression in many works.
2	A small amount of work has been produced which demonstrates a limited understanding of the conceptual and/or technical underpinnings of artistic expression. The work illustrates a limited or diffused exploration of ideas appropriate to the visual arts, and a limited ability to resolve concept, media, and technical expression.
1	An inadequate amount of work has been completed, and is lacking in evidence of technical skill or relevant knowledge of artistic expression.

Assessment reliability has been achieved in qualitative assessment when multiple judges are able to independently assign equivalent value to the same work.

Benchmarking is a common practice used to clarify the meaning of rubric statements defining performance levels. This is done by selecting student work samples that exemplify the meaning of each level of the rubric. For the most effective results, it is advisable to have a group of selected expert judges discuss the rubric and agree upon best examples to use as benchmarks for each level.

Content Standards are a written description of what students should know and/or be able to do in a particular content or subject area. The expectations articulated in the content standard outline the knowledge, skills, and abilities for all students in the subject area.

Performance (Achievement) Standards express the degree to which students have achieved mastery of the content standards. Such statements are usually expressed using criteria and related rubrics that describe levels of performance from unsatisfactory through acceptable to high levels of performance. Performance Assessment tasks are often complex and may include such activities as the creation of a portfolio, reflective diaries, a performance, a presentation and so on. This concept is closely related to authentic assessment.

REFERENCES

The majority of the above material is sourced from the following encyclopedia entry:

Boughton, Douglas G. (In Press) Assessment of Student Art Learning. *In International Encyclopedia of Art and Design Education*, Editor Kerry J Freedman, (Vol. 2) Hoboken, NJ: John Wiley & Sons, Inc.

Other references cited above:

Boughton, Douglas, Eisner, Elliot, Ligtoet, Johan, eds. 1996. *Evaluating and Assessing the Visual Arts in Education*. New York, NY: Teachers College Press.

Eisner, Elliot W. 1996. "Evaluating the Teaching of Art" in *Evaluating and Assessing the Visual Arts in Education*. Edited by Boughton, Douglas, Eisner, Elliot and Ligtoet, Johan. 75-94. New York, NY: Teachers College Press.