chapter 17

Assessment in the Secondary Music Classroom

Martin Fautley and Richard Colwell

The issue of assessment in music education in the secondary school is one of concern in a range of contexts, including teaching, learning, accountability, policy, and politics. In order to investigate assessment in the secondary school, we need to begin by understanding what assessment is, what the terminologies involved mean, what the implications of assessment are for learners, teachers, program organizers, administrators, legislators, and other interested stakeholders, and what constitutes secondary school music.

In the course of this chapter we will consider the following issues in student assessment (program assessment is not considered):

The context of assessment

* Uses and purposes of assessment
* Legitimizing assessment

Assessment and music pedagogy

The Context of Assessment

Secondary school music experiences can be as broad as the definition of music. Even in those countries with a secondary school music requirement, students can usually elect the musical experience of interest. Thus, any portrayal of assessment practices will be as wide-ranging as are the curricular offerings. We discuss assessment which happens for pupils aged between roughly 11 and 18 years old. Between these two age delimiters lie a wide range of experiences and musical encounters, from formal compulsory music lessons for all, through various forms of optional take-up of music, including a range of electives, through to systems where music itself is optional; within all contexts there is also a wide range of musical learning that takes place outside formal curricula. The content component of music courses also varies widely, from contexts where there is a considerable statutory program of study, such as a national curriculum, through to those where the main delivery of music lies in participation in musical ensembles (content standards). Assessment, however, focuses on performance standards (student competency on the content standard). We can be most definite about assessment when a purpose of secondary school education is preparation for college, specifically a career in music. The types of assessment for students electing this track are similar, whether Advanced Placement music theory examinations in the United States, the history and theory exams in Israel that follow a sequential curriculum, or the performance and theory examinations in Taiwan. When the state is involved in these examinations, scores on these examinations can determine scholarship and admission decisions.

Assessment involves making judgments, and is closely tied to, or synonymous with, issues of evaluation. Assessment has a long history in music; for example, singing and playing contests. The secondary school context of music instruction also needs explaining. There are a range of approaches to secondary schooling across many national contexts, and some of these are dissimilar from each other.

Assessment and Standards

Assessment consists of designing, conducting, interpreting, and communicating the results of an investigation into learning. One assigns value, one describes the meaning of data and observations, one synthesizes experiences, and the resulting judgments indicate the merit, worth, and significance of the educational venture. Assessments can be influential in determining support for or reduction of programs, they can aid in establishing priorities and instructional emphases, and they can be specific or general. In the USA, and, increasingly, internationally too, the AERA publication *Standards for educational and psychological testing* (AERA, 2014) provides much useful information for this aspect of the work of both test-setters, and classroom teachers more generally.

The term “high-stakes” assessment is given to summative assessments that certificate significant aspects of attainment. These are the ones that give rise to publicly accountable assessments, in the form of end-of-course examinations, school leaving certificates, degree classifications, and so on. They are considered high-stakes as the implications of passing them can be significant, and the implications for failing them can be an impediment.

There is no doubt that the international standards movement in curriculum and testing in a few core subjects, such as mathematics, has also influenced the assessment movement in music. State agencies and professional organizations are exploring the development of music assessments modeled on the core assessments used for national and international comparisons. These efforts are valuable and of interest, but high-stakes tests in music are not likely to be required. For example, one issue in assessing is determining what level of competency constitutes “proficient.” Australians (New South Wales) who had an assessment mandate have experimented with use of a cutoff point between proficient and not proficient, as well as scores indicating other levels of proficiency. Establishing these cut-points is important. The standards movement consists of three standards: content standards (the appropriate level of the content being presented); performance standards (whether the level of performance, knowledge, understanding, etc. is at the expected age level); and opportunity-to-learn standards (have the students been provided with adequate resources, expert teachers, and instructional time) to reasonably be expected to attain the content and performance standards. These standards may differ depending on mandates. In Mexico, outcomes are expected to include awareness, contemplation, expression, appreciation, contextualization, and creation. Caribbean nations have agreed on listening and appraising, performing, and composing. Belarus and Russia seem to assess music instruction based on observation, singing and playing of instruments, creative assignments, oral responses, essays, projects, and listening exams that test composer and/or selection recognition. Taiwan has established cognitive objectives as being 25% of any assessment, psychomotor 50%, and affective, which includes participation, appreciation, and creating, 25% (Holmes et al., 2010). In the United States, percentages are not specified but include performing, creating, and responding. Categories or mandates in other countries suggest composing (including arranging), audiation, performing, literature studies, skills, and aesthetics—musical styles and idioms.

Uses and Purposes of Assessment

Assessment and Testing

In some quarters there is an assumption that assessment is primarily concerned with testing. This assumption is breaking down in contemporary assessment discourse. It is the case that testing involves assessment, but one can measure without assessing—assessment requires a judgment about value that a score on a test does not provide. It is entirely possible to make an assessment judgment concerning a pupil’s music-making that does not involve a test.

The construction of all assessment strategies requires consideration of a table of specifications—what aspects of music instruction should be evaluated and their relative importance. The division between cognitive, psychomotor, and affective objectives is a type of table of specifications, although a table often consists of multiple specific topics and weightings. Usually, consideration has also been given to a taxonomy of learning to ensure assessing a wide range of competencies from simple to complex, from basic factual knowledge to competence in use of that knowledge in analyzing, synthesizing, and evaluating.

Assessment and Evaluation

A distinction needs to be drawn between assessment and evaluation due to country differences. For educational purposes, however, in the United States, evaluation may have an empirical aspect (that something is the case), or normative aspect, where one determines the value of something, as used commonly in program evaluation; whereas in the United Kingdom, evaluation usually means “. . . the process of collecting evidence and making judgments about programs, systems, materials, procedures and processes; ‘assessment’ refers to the process of collecting evidence and making judgments relating to outcomes, such as student’s achievement” (Harlen, 2007, p. 12). Likewise, in the UK the term *teacher evaluation* normally refers to evaluation *by* teachers, whereas in the US it means evaluation *of* teachers and teaching (Kane *et al.*, 2014). In Spain, assessment and evaluation are interchangeable but have two meanings: assessment can mean grading, as assigning a formal grade for school reports and promotion, or it can have the broader meaning described in this chapter. In discussing student achievement in the United States, the terms “assessment” and “evaluation” tend to be interchangeable. Measurement, however, is usually defined as the use of tests designed to report any change in behaviors by means of assigning categories or numbers to these, which result in a score that can then be used in summative assessment. Measurement tries to be objective; assessment/evaluation tends to be more subjective. We shall not attempt to resolve in this chapter the definition differences in various cultures.

Summative Assessment

In many national contexts summative assessment is what is meant when assessment is being discussed; in other words, for some teachers summative assessment is assessment!

Summative assessment takes place at specific points in a course of study, and is designed to sum up the attainment of the learner. Summative assessments are often put in place at the end of instructional units, or at the end of a course, especially when they provide a certificatory function. Thus a graduate will talk of having a first class degree, or an upper second, and a pupil will talk of having grade 5 flute. Summative assessments include recitals, auditions, contests as well as formal, scheduled, assessments.

Summative assessment can take place on a smaller scale too. Teachers who give their class a test may be employing summative assessment. The pupil who gets a score of 9 out of 10 will be considered to have done better at the test than the pupil who gets 4 out of 10. These small-scale classroom tests can be used to build up a picture of attainment over time. Summative assessment is common in accountability contexts and when grades in music are customary or required. These assessments, auditions, the research in New South Wales, and Item Response Theory (IRT; uses item analysis and a statistical program to estimate that if a student knows the answer to question X that s/he also knows the answer to question Y) work in Germany are all summative examples. Summative assessment is used for accountability and for making policy decisions. Summative assessments are criterion-referenced measures; the criterion can be a cut score, a rich description of what musical competencies indicate “proficient” on a score of three or above or an Advanced Placement examination. With criterion-referenced measures, all students can be successful. The results of summative assessment should result in reflection by both pupil and teacher: What can I do differently next time to improve? Why did the students not learn what was emphasized?

Formative Assessment

Studies have suggested that it is formative assessment that makes the most difference to student learning. Black and Wiliam found that the differences it makes are “quite considerable, and [are], amongst the largest ever reported for educational interventions” (1998, p. 61). Black and Wiliam’s data demonstrated the effect size of proper use of formative assessment—effect size indicating the degree of student improvement. Proper use is key; large numbers of studies of the use of formative assessment showed no or even negative impact. Effect size is similar to the normal curve; an effect size of .42 is equivalent to a percentile gain of 16 percentage points, for example a student’s score could increase from the 50th percentile to the 66th percentile; in other words, a student’s score could increase from 5 to 6.6 out of 10. A meta-analysis by Hattie (2009) of 15 studies found effect sizes ranging from .23 to 1.12. In music education formative assessment has a long history of being a part of the daily work of the classroom and instrumental teacher. Formative assessment procedures involve working with students so that they know what it is they need to do in order to improve. Validity and reliability data are not crucial with formative assessment—excellent teachers, however, have valid content objectives and tend to be consistent in their interaction with students.

One of the most widely misunderstood aspects of assessment currently is the distinction between formative and summative assessment. This has been exacerbated by these words being interpreted in different ways in different national contexts. Formative assessment is concerned with taking learning forward. It is not one or multiple tests; true formative assessment is a process. Music teachers in all national contexts will have incorporated aspects of formative assessment into their teaching, possibly without realizing that is what they are doing! Formative assessment normally involves dialogue; it is not “top-down” assessing. Here are some examples:

Instrumental lesson: when a learner plays her piece to the teacher, and the teacher makes comments, s/he will talk with the learner, to help the learner understand what it is that needs to be done, and to help clarify issues and misconceptions.

* Classroom composing: when a group of pupils compose using classroom instruments, the teacher will discuss with the learners how their work could be improved and what the pupils’ understanding of the task involves, and will help the pupils identify what could be done next.
* Choir rehearsal: when a choir is having sectional rehearsals, the leader makes observations as to what could be done to improve, gets the singers to try out alternatives, and works with them to develop the aspects in question.

Ensemble rehearsal: when the second clarinets play an eight-bar phrase and the director points out rhythmic inaccuracies, asks the players to try again, and then checks that they understand the differences by discussing with them and possibly doing a before/after comparison.

As teachers make constructive comments and observations, as well as modeling, they are undertaking formative assessment. What this requires from the teacher is that s/he needs to know what it is that needs to be addressed, what the next sequential (more challenging) steps to be taken involve, and then shares these, usually orally, with the learners. In assessment literature this is characterized as feedback. In music, oral/demonstration feedback is a normal feature of learning. Oral feedback is spontaneous, and located in the immediacy of a learning episode. Written feedback has a part to play, too, but the immediacy of the oral is a key feature of formative assessment. Viewed from this perspective, it is hard to imagine teaching without assessing; indeed, as Swanwick rightly observes, “. . . to teach is to assess” (1988, p. 149). What this means is that as true formative assessment is a process, it needs to concern itself with what is to be done next in order for the learner to improve. This is assessment that is done by the teacher with the pupils. Formative assessment does not in and of itself yield grades or levels, nor does it compare learners one with another. It is individualized and personalized, and concerned with taking a specific learner or group of learners forward in their own music-making. To this extent some teachers have observed that it does not feel like assessment at all! For example, in Rusinek’s study of disaffected learners in Spain (2008), the teacher used formative assessment strategies (although they are not referred to as such) to engage students who were otherwise at risk of failure in the school system. Classroom assessment is not (necessarily) formative assessment; it must lead to improved instruction (Black & Wiliam, 2004).

Characteristics of Effective Formative Assessment

It is a goal of all music teachers to improve the learning and performance of their students. There are many ways by which this can be undertaken in music education, but it is likely to be formative assessment which is at the heart of these. There are a number of features that formative assessment will take:

It will be dynamic: Students will play their music, and the teacher will work with the pupils to explain, probably modeling, what it is that could be done to improve.

It will be a part of the learning encounter: Formative assessment does not take place separate from teaching, it is embedded within it.

It will involve feedback: Teachers will talk with, not at, their pupils about the learning. This enables the teacher to understand the pupils’ context, clear up misunderstandings, and produce a personalized response based on the music they have just heard, not simply general exhortations to do better.

It will be reactive: As a result of a learning episode, the teacher needs to be prepared to change aspects of the pedagogic approach. This can mean choosing different pieces to perform, choosing specific studies to embed technique, closing gaps in prior learning through careful review prior to initiating new tasks. It means the teaching reacts to the learners; it does not carry on regardless.

Students know what it means to get better: Formative assessment encounters based on the above tenets should clarify what the next steps are that the students need to take, specific to their needs. This involves delineating clearly what comes next. To do this the teacher needs to share their understandings of quality outcomes with the learners; again modeling can help. Students can begin to take charge of their own learning when they have the competence to “validly” assess their own learning, and, where appropriate, that of the ensemble.

The Formative Use of Summative Assessment

Key to understanding the differences between formative and summative assessments is to realize that, stripped of any context- or nation-specific overtones, formative assessment is about improving learning; summative assessment is about auditing it. One of the reasons that this distinction has not been widely grasped is that much of what passes for formative assessment is not formative assessment per se, but is summative assessment used in a developmental fashion.

Some national and local contexts involve the regular use of testing. Such testing, as we have seen, involves summative assessment. Telling a student that she has achieved, say, 7 out of 10, level 5, or a score of 27, does not tell the learner what it is she needs to do in order to improve. The interpretation of test scores in a way that does help the students improve is known as the formative use of summative assessment. Figure 4.4.1 (from Fautley & Savage, 2008, p. 27) gives a visual representation of this.

Figure 4.4.1 shows that the formative use of summative assessment is the way teachers can use the results from tests, working in ways we discussed in the section on formative assessment above, to help their learners move to the next level of attainment. This needs to be interactive, as simply knowing that a mark of 6 out of 10 was achieved does not tell the student how to achieve 7 out of 10. This is because “marks, levels, judgmental comments or the setting of targets, cannot, on their own, be formative. Pupils may need help to know how they can improve” (Mansell et al., 2009, p. 10).

The differences between formative and summative assessment were summarized by Mansell et al. (2009, p. 9):

Summative comes at the end of learning episodes, whereas formative is built in to the learning process.

* Summative aims to assess knowledge and understanding at a given point in time, whereas formative aims to develop it.
* Summative is static and one-way (usually the teacher or examiner judges the pupil), whereas formative is ongoing and dynamic.

Summative follows a set of predefined questions, whereas formative follows the flow of spontaneous dialogue and interaction, where one action builds on (is contingent on) an earlier one.

**Figure 4.4.1**: Assessment modalities. From Fautley and Savage (2008, p. 27).

Self- and Peer Assessment

Self- and peer assessment are key assessment strategies. The ideal situation is one where the student comes to hold the same concept of quality as the teacher, and knows what to do in order to develop his or her own learning, or move to the next step. In order to do this, both summative and formative assessment require a clear and concise conceptualization of what is to be learned. Formative assessment requires that the teaching and the sequencing of understanding be sequential, the sequence understood by pupil and teacher, and that it be based on some type of taxonomy that leads from knowledge of terminology to higher order thinking.

Self-assessment is often undertaken by pupils in terms of what they wish to be able to do next, and based on their own realistic understandings of what they can do. This is a point outlined by Alda de Oliveira in her discussion of the music education of street children in Brazil (2000).

Portfolio Assessment

The notion of portfolio assessment is a key issue in educational thinking today, and has a number of benefits for the music educator. Portfolio assessment involves building up evidence over a range of attainments as evidenced by the growth of the individual. For example, in contexts where performing is central to curricular music, one solution in secondary music ensemble experience, rather than giving each member a grade, is to rely on a student portfolio. Each spring, or more often, the ensemble presents itself to the public and to school administrators through a public performance of works mastered during the year. On request, the ensemble could sight-read a number, and/or students could demonstrate their understanding of the music performed by responding to questions from the audience. Small ensembles and soloists could perform as well. The material in a portfolio still must be assessed either by a range of descriptors or by gain scores. The interpretation could be pass-fail or a multiple point continuum. The portfolio is a type of summative assessment.

For more general classroom music, portfolios would normally be expected to include a range of composing, performing, and listening activities. In Taiwan, Ming-Jen Chuang has described their utilization in schools by a group of teachers, observing that the range of responses they produced enabled the teachers concerned to become more reflective in their work (Chuang, 2007). For twenty-first-century purposes, the use of e-portfolios readily allows recordings of student work to be located centrally, along with supporting documentation, and any scores or performing instructions. E-portfolios permit a wide range of material to be included, including the all-important sonic representations of music composed and performed by the student concerned.

Legitimizing Assessment

All assessments must be fair. In addition, they must be valid and reliable.

Reliability

Reliability is concerned primarily with the consistency of the assessment. For example:

If a student undertook the same assessment on two subsequent occasions, would the same results be obtained by that student?

* If a single assessment by a single student is graded by two different assessors, will the results be the same?

If the same assessment was undertaken by two groups of students of similar ability, would both groups of pupils achieve the same results?

High-stakes testing is normally criterion-referenced, and reliability should be quite high.

Validity

Validity refers to how any data are used to arrive at the truth, and the interpretation of these data. Thus, there are multiple descriptions (types) of validity, and utilization of them is going to depend to some extent on the context, as well as the uses and purposes of the assessment. Four types of validity commonly encountered in the literature are:

Face validity: This is where an assessment looks as if is assessing that which it is supposed to be assessing.

* Content validity: Here the data need to be relevant to the subject matter.
* Construct validity: A construct in education refers to an abstract item, such as pitch, or rhythm, which should be understood and uniquely definable.

Consequential validity: Refers to the consequences that assessment and testing can have on learners, teachers, systems, and society (for a detailed discussion of validity issues in assessment, see Kane, 2006).

There are trade-offs between reliability and validity:

 . . . attempts to increase reliability which generally means closer and closer specification, and use of methods that have the least error. It [attempting to increase reliability] results in gathering and using a restricted range of evidence, leading to a reduction in validity. On the other hand, if validity is increased by extending the range of the assessment to include outcomes such as higher level thinking skills, then reliability is likely to fall, since many of these aspects of attainment are not easily assessed. (Harlen, 2005, p. 247)

Validity also needs to be considered with regard to the context in which an assessment has occurred. So, for example, a single outcome score might superficially be the same, for example 58%, but this does not imply that it will have a transferable meaning from one context to another. 58% in the context of a band performance means something very different to 58% in the context of a harmony examination. The notion of context needs to be considered carefully here (Colwell, 2016).

Objectivity and Subjectivity

Reliability and validity are important issues for music education, because in an attempt to increase reliability, it is all too easy to fall back on things that are easily assessable, which, while probably reliable, are not necessarily valid in measuring aspects of musical learning. For example, can an evaluation of an advanced-level pianist’s interpretation of a Beethoven piano sonata be undertaken solely by assessing his or her fingering? This is fairly straightforward to assess, but probably does not vary much between performances, and so is reliable, but as an assessment of musical worth, its content validity is questionable. Not dissimilar instances are regularly to be found in classroom and instrumental music lessons.

In music some questions have right answers, such as when was Mozart born, and some are more open to interpretation, such as in a specific piece of harmony, is IIb–V–I a better cadential progression than IV–V–I? Extrapolated further, this means that issues of objectivity and subjectivity will inevitably form part of the thinking concerning assessments, and the purposes to which they are put. There is a tendency to treat assessment data as being only worthwhile when it is entirely objective, but this is problematic. Presently there seems to be an emphasis on objectivity in education, whereas in industry, employees are observed, and subjectivity predominates. Indeed, as Gipps warns, “assessment is not an exact science, and we must stop presenting it as such” (1994, p. 167). There will be many areas of subjective assessments made in music education, and this is to be expected. What is appropriate, however, is for assessments to be as reliable and valid as they can be. But this in itself raises another issue for music education, as assessing things that are easily assessable may not result in assessing the things that should be assessed, the unfortunate consequence of which is that things come to be assessed because they can be. This can lead to situations where those things that are assessed are technical or organizational matters of music-making that are easy to assess, for example a composition either does or does not contain an ostinato, whereas assessing creative thinking involved in a composition or whether the results of a performance can be considered to be musical or not very musical are much more difficult judgments to make.

Contained within issues of objectivity are those that appertain to assessment of different standards of proficiency. Faced with a bewildering range of contemporary styles and types of music produced by secondary-school-age students, assessment strategies need to keep pace with this. For teachers who are “insiders” in the music in question, this may not be too much of a problem; for “outsiders,” the task can seem daunting. What is meant by a musical performance of a Chopin etude? What is meant by a musical performance of a blues piece? What is meant by a musical performance of thrash metal? What is a musical composition for a 12-year-old pupil using a Glockenspiel? Are all these examples evidencing the same notion of musicality? Do they compare? What if they are all by the same pupil? Also relevant here is the notion of consensual assessment (Amabile, 1982), where judges with appropriate expertise formulate both gradings and criteria, making criterion-referenced validity of prime importance. Where issues of validity and reliability will arise is when assessments of different types of musical attainment are compared one with another, and this returns us to notions of objectivity. For example, how might the grade given to a rock band be equivalent to that given to a baroque recorder quartet? Should it?

Assessment Data

Assessment data, formative and summative, refer to that which arises when assessment is undertaken. Assessment data can be both tangible and intangible, taking a number of forms, including recordings, conversations, test scores, pupil writing, pupil performances, musical scores, and written comments. So far it is the purposes of assessment that have been discussed. However, the uses to which assessment data are put are significant, too, especially as the information that arises from assessment sometimes has to serve a range of uses. Formative assessment has one clear aim, which is to improve teaching and student learning. In the case of summative assessment, however, there can be a range of uses to which assessment data are put.

Using Data from Assessments

The most important and the most controversial component of any type of assessment is how the data are used. What does a score of 88/100 on a test tell us about the student or the program? or a score of II from a music festival or contest, or a “satisfactory” on a student composition? Interpreting data from any type of assessment requires knowledge, experience, and a full understanding of the characteristics of tests, testing, and students. A student’s background, his or her talent, socioeconomic level, home and community support, level of effort, and much more, including tradition and expectations of the community, are variables that contribute to the controversies of assessment and the level at which it has meaning for music teaching and learning. Assessment of secondary school music in all cultures consists of two major types of assessment, formative and summative. Either type of assessment begins with the statement of a clear goal of what it is to be assessed, and those being assessed must understand the goal or goals. Without this clear statement, the assessment will lack reliability and validity. A survey of assessment practices in Sweden found that students seldom knew the purpose of the assessment, did not receive feedback from the assessment, and did not know whether the results meant that they were performing at the expected level of competence (Murphy, 2007).

Designing assessments that have to meet the requirements of different groups of users can distort the assessment being undertaken. Take the example of a teacher who wants to see how well her class is able to perform a specified piece of music. She may well set up a performing situation where the pupils play the piece, and the teacher grades them. The results from this can be used for the teacher to decide what is needed to do next in order for her learners to improve, and for the learners to know how well they performed. However, if the same sets of data are also used to measure how well that teacher has worked at teaching her class the piece of music, then the teacher is likely to approach the teaching in a rather different fashion, ensuring that the focus remains on the end product at all times, and avoiding any important incidental learning that could take place along the way. What is happening here is that assessment data are having to simultaneously fulfill more than one purpose, what Boud (2000) refers to as the “double-duty” of assessment. Using assessment data for purposes for which they were not intended might not be helpful; “clarity about the legitimate and illegitimate uses of assessment data, and the intended or unintended consequences of those uses, is crucial” (Mansell et al., 2009, p. 8).

This can be taken a stage further, too. If assessment results are going to be published nationally and enable comparison of a specific school with other schools, then the management team of the school is also likely to want to monitor the way the music teacher has taught the class the performance piece, in order to maximize the school’s position. In some countries, publication of such grades enables newspapers to print “league tables” of schools classified by the results from summative assessments. This aspect of assessment raises issues of accountability, and assessment data used in this way can cause problems for music teachers. Indeed, there is a danger that in some countries externalized assessment becomes a means whereby the content can be determined centrally: “State standards reflect value choices about what is most important for students to learn and what constitutes mastery of that knowledge. But different constituencies have different ideas . . .” (Colwell, 2007, p. 6). This is an important observation, and is key to understanding different local and national contexts concerning assessment. What this can result in is known as “teaching to the test,” and occurs where high-stakes assessment systems place considerable weight on assessment results. This can result in a narrowing of the curriculum, and of learning opportunities, as teaching becomes focused solely on final assessment; this is known as “assessment backwash.”

Many jurisdictions are working on ways in which data generation and collection by teachers can be improved (Brookhart, 2013; 2014; 2015; 2016; (Gardner, 2012) . As the performativity role of assessment data (Ball, 2003, Wiliam, 2011) becomes ever more tangible, and the role of the OECD PISA (programme for international student assessment)(PISA website) assumes greater international importance for comparative purposes, these various assessment systems are increasingly likely to become less disparate, and more likely to converge.

Assessment and Music Pedagogy

Knowledge Types in Music Education

In designing assessments, the first thing that needs to be done is for the assessment designer to decide what it is that will be assessed and ensure that the objective/task is clearly understood by all students. In music education there is a long history of a split between learning and doing. This needs some consideration, as assessment of an activity will derive from task criteria, whereas assessment of learning will derive from learning criteria, and it is likely to be the case that these are different. Teachers often have finely honed skills in deriving task criteria from action; what can be less clear is the specification of learning criteria. The implications of this for practice can be considered through the example of classroom performing. If a class is engaged in a performance using tuned and untuned percussion instruments, say, then an individual’s contribution to the performance of the piece of music can be assessed by such factors as being able to play correct notes, on the beat, and with a reasonable technique. But what have they learned by doing this? Have they learned about playing on the beat? Is this assessed in performance? If they have learned it already, why are they doing it again? In other words, how is assessment of learning distinguished from assessment of doing? This key question takes us to the heart of the matter in music education assessment. An answer to this question can be found in a consideration of knowledge types involved.

Ryle (1949) wrote of “knowing that” and “knowing how.” These two types of knowledge also equate with the notion of declarative knowledge and procedural knowledge. In the United States, Bennett Reimer (1992) has suggested that four types of knowledge are relevant to musical knowing -- knowing within and knowing how as primary, knowing about and knowing why as contributory to the first two. Knowing within is least amenable to verbal description, but “feeling within” is apt. These types can be described in terms of objectives; to perceive, discriminate, feel, and evaluate works; to be aware of historical, social, cultural, political, and religious contexts, and to be cognizant of the issues that surround them, thus influencing perceiving, understanding, creating, and judging. What classroom assessment will need to do is to focus attention on the knowledge type that is most appropriate.

Devising Assessment Criteria

Music education has a long history of devising assessment criteria for performing music. In the United Kingdom, the Associated Board of the Royal Schools of Music has been devising such assessments for more than a century, and similar principles of a graded sequence of examinations where content is progressively more challenging are well embedded in a range of national contexts. However, devising assessment criteria for other areas of classroom music, such as composing, tend to be less well developed. In developing assessment criteria for music education a number of points need to be borne in mind. These include:

A criterion should have a degree of exclusivity, in that it should normally be specified closely enough to be able to evaluate a single item, skill, or construct without too many extraneous variables; in other words, it helps if a criterion can relate to a singularity.

* A criterion should be assessable, in that it should be possible to ascribe a hierarchy of attainment.
* A criterion should have some relationship to the whole, in that it should not be evaluating a tangential aspect of attainment.
* A series of criteria that endeavor to deconstruct a whole should, when taken together, go some way toward formation of an overall impression of the whole.

That which is deemed important, but hard to assess, should not be omitted at the expense of that which is easy to assess, but less important.

Assessment of Performing

The application of assessment criteria as outlined above finds its most ready expression in performing music. In some national contexts, performing music is almost the only one in which music education takes place. In others, performing is part of a more general program of music education. Whatever the system, as with other aspects of musical learning, teachers will need to consider formative and summative aspects of performing separately.

According to McPherson and Schubert (2004, p. 64) there are at least four main types of competences that are addressed using assessment criteria in music: technique, interpretation, expression, and communication. These are areas that the teacher will want to assess formatively in order to affect improvement and summatively in order to measure and audit.

In the case of contexts where assessment of solo and ensemble performance is deemed appropriate, there are a number of complications. A form of checklist is common that has been proved to be fairly reliable but with little evidence of validity. A vocal performance could be rated on tonal, rhythm, executive technique, and expression, each with subcategories. For example, executive technique may be remarkable because of student posture, diction, vibrato, focused and centered tone quality, and uniform vowel pronunciation. A checklist furnished to judges of ensembles (the checklist necessary to increase reliability) may include performer controlled elements of tone (natural, free, vibrant) dynamics, breath/stick/bow management, pitch, and rhythm; each weighted as to its importance, accompanied by an evaluation of the conductor’s interpretation, musicianship, technique, and his/her management of stage deportment, again each weighted, with deportment carrying the least overall weight in a total score.

Assessment of Composing

Many national music education contexts involve composing. This enables students to actively connect with music-making. It does not normally involve learners in staff notation writing, but is concerned with creating music directly using classroom instruments or information and communication technology (ICT), and is often undertaken as a conjoint group activity. Composing is a creative act, and creativity is notoriously difficult to assess. What the assessment of composing will entail are decisions made by the teacher as to the division between assessment of composing process and assessment of compositional product. Assessment of product can normally be undertaken using teacher or externally devised criteria, and will entail summative assessment strategies. Assessment of composing as process will often entail formative assessment. This means that the teacher needs to enter into dialogue with the students with regard to the music they are producing, using the formative strategies outlined above. (It is important to note that assessment of composing needs to be separated from that of performing; it is the composing aspect that should be to the fore here.)

Composing in general music classes often entails a series of stages, including generative, revising, and organizational. The role of the teacher in formative assessment is to help the learners work their way through these stages, aiming at helping them realize their ideas. Intentionality can be a key factor, which is why formative assessment strategies, with their emphasis on discussion and feedback, are important.

In a study of the criteria used for assessing composing in Australia, Beston found that teachers tended to utilize composition-specific criteria wherein “each composition generates specific groups of criteria which centre predominantly on applications of musical concepts” (2004, p. 37). It was also found that “teachers avoid criteria in their judgments which are less quantifiable or more challenging, such as originality” (p. 38). In Sweden, in a study of composing using ICT (Nilsson & Folkestad, 2005), it was found that it was through discourse that young people’s creativity could best be understood. This again emphasizes the key role that formative assessment plays in taking learning forwards.

Assessing Informal Music Learning

With the rise of informal musical learning, such as the Musical Futures program in the United Kingdom (www.musicalfutures.org.uk), new demands are being made on teachers with regard to assessing informal learning. What will be of importance here is the notion of assessing development, and this is again process-based. In many ways this has links with the way composing is assessed, in that intentionality is again key. Here the role of the teacher is to help the learners with understanding what it is they wish to achieve, and then trying to remove barriers that prevent them from achieving it. This entails knowing what the learners wish to achieve, again a formative assessment judgment.

Conclusion

Assessment in music education is a complex and contested area. There are no easy answers. Assessment of musicality, of musical performances, of composing, of improvising all bring their own challenges. In this chapter we have only begun to scratch the surface of the topic. For many teachers, working in situations where performativity and accountability are the watchword, the role of assessment in day-to-day teaching and learning has assumed significance. Both summative and formative assessment are essential in secondary school music, and both must be done well. We have pointed out the dangers of an overreliance on assessment assumptions, and have made suggestions as to how assessment can be used for the purpose that music teachers are likely to be most focused on in their everyday working: improving the learning of all pupils.

Reflective Questions

How do you employ formative assessment in your teaching?

* How do your pupils know what to do to improve?
* Who are the results of assessments in music education aimed at? You? The pupils? Parents? School management teams? District auditors?
* Why do you teach what you do? Does assessment inform it? If so, how?
* What use do you make of tests and test scores? Why?

Do you have a clear idea as to what improvement in what you teach entails? How do you articulate this to the learners?

Key Resources

Brophy, T. S. (2000). Assessing the developing child musician: A guide for general music teachers. Chicago: GIA Publications.

Brophy, T. S. (Ed.) (2008). Assessment in music education. Proceedings of the 2007 symposium on assessment in music education, Florida. Chicago: GIA Publications.

Colwell, R. (2002). Assessment’s potential in music education. In R. Colwell & C. P. Richardson (Eds.), The new handbook of research on music teaching and learning: A project of the Music Educators National Conference (pp. 1120–1158). Oxford, New York: Oxford University Press.

Fautley, M. (2010). Assessment in music education. Oxford: Oxford University Press.

References

AERA, A., NCME, 2014. *Standards for educational and psychological testing,* Washington, DC, American Educational Research Association.

Amabile, T. (1982). Social psychology of creativity: A consensual assessment technique. Journal of Personality and Social Psychology, 43, 997–1013.

Beston, P. (2004). Senior student composition: An investigation of criteria used in assessments by New South Wales secondary school music teachers. Research Studies in Music Education, 22(1), 28–41.

BALL, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy,* 18**,** 215-28.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5(1), 68.

Black, P., & Wiliam, D. (2004). Classroom assessment is not (necessarily) formative assessment (and vice-versa). Towards coherence between classroom assessment and accountability (pp. 183–188). 103rd Yearbook of the National Society for the Study of Education. part II, Chicago: University of Chicago Press.

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. Studies in Continuing Education, 22(2), 151–67.

Chuang, M.-J. (2007). A class portfolio can be feasible in general music classes. In T. S. Brophy (Ed.), Assessment in music education. Proceedings of the 2007 symposium on assessment in music education. University of Florida (pp. 259–272). Chicago: GIA Publications.

Colwell, R. (2007). Music assessment in an increasingly politicized, accountability-driven educational environment. In T. S. Brophy (Ed.), Assessment in music education. Proceedings of the 2007 symposium on assessment in music education. University of Florida (pp. 3–16). Chicago: GIA Publications.

Colwell, R.  (2016). Tergiversation Today: Interpreting Validity.  In T. S. Brophy, J. Marlatt, & G. Ritcher (Eds.), *Connecting practice, measurement, and evaluation.* Chicago: GIA Publications.

de Oliveira, A. (2000). Street kids in Brazil and the concept of teaching structures. International Journal of Music Education, 35(1), 29–34.

Fautley, M., & Savage, J. (2008). Assessment for learning and teaching in secondary schools. Exeter, UK: Learning Matters.

GARDNER, J. (Ed.) 2012. *Assessment and learning (2nd Edition),* London: SAGE.

Gipps, C. (1994). Beyond testing: Towards a theory of educational assessment. London: Falmer Press.

Harlen, W. (2005). Trusting teachers’ judgement: Research evidence of the reliability and validity of teachers’ assessment used for summative purposes. Research Papers in Education, 20(3), 245–270.

Harlen, W. (2007). Assessment of learning. London: Sage.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses related to achievement. New York: Routledge.

Holmes, A. V., Gokturk, D., Aguilar, B. E., & Chen, J.-J. (2010). Assessment “over the ocean”—outside the US. In T. S. Brophy (Ed.), The practice of assessment in music education: Frameworks, models, and designs (pp. 95–102). Chicago: GIA Publications.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed.) (pp. 17–64). Westport, CT: American Council on Education & Praeger Publishing.

Kane, T.J., Kerr, K.A., & Pianta, R.C. (Eds.) (2014*). Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. San Francisco: Jossey-Bass

Mansell, W., James, M., & Assessment Reform Group. (2009). Assessment in schools. Fit for purpose? A Commentary by the Teaching and Learning Research Programme. London: Economic and Social Research Council, Teaching and Learning Research Programme.

McPherson, G., & Schubert, E. (2004). Measuring performance enhancement in music. In A. Williamon (Ed.), Musical excellence: Strategies and techniques to enhance performance (pp. 61–84). Oxford: Oxford University Press.

Murphy, R. (2007). Harmonizing assessment and music in the classroom. In L. Bresler (Ed.), International handbook of research in arts education (pp. 361–388). Dordrecht, Netherlands: Springer.

Nilsson, B., & Folkestad, G. (2005). Children’s practice of computer-based composition. Music Education Research, 7(1), 21–37.

PISA website: https://www.oecd.org/pisa/ Accessed 10/11

Reimer, B. (1992). What knowledge is of most worth in the arts? In B. Reimer, R. Smith, & K. Rehage (Eds.), The arts, education, and aesthetic knowing: Ninety-first yearbook of the National Society for the Study of Education (pp. 20–50). Chicago: National Society for the Study of Education.

Rusinek, G. (2008). Disaffected learners and school musical culture: An opportunity for inclusion. Research Studies in Music Education, 30(1), 9–23.

Ryle, G. (1949). The concept of mind. Harmondsworth, UK: Penguin Books.

Swanwick, K. (1988). Music, mind, and education. London: Routledge.

WILIAM, D. (2011). *Embedded formative assessment.* Bloomington, Indiana, Solution Tree Press.