

Whose performance is it anyway? Reflections on examining music, meanings, standards, and reliability in an international context.

Paper to be presented at the 33rd Annual IAEA Conference, Baku, September 2007

Lead author and correspondence to:

Dr. Jonathan H Robbins
The Talent Centre Nottingham Weymouth Dorset DT3 4BH England
+44 (0)1305 814820 www.talent-centre.com

Second author:

Professor John Howard Thames Valley University, London College of Music Examinations

Presenter

Dr. Jonathan H Robbins

Abstract

Forms of graded examinations in music developed in England over at least 150 years are being used internationally as qualifications that recognise achievement and standards of performance. An overview of the history of these examinations and their use internationally is provided. The role of the teacher in preparing candidates for these examinations is described. Matters of interpretation, phrasing, and musical understanding by candidates in the context of assessment are discussed. Consideration is given to the influence of the teacher on the examination result for a candidate, together with that of the external examiner as well as the meaning of standards and reliability in international settings. Applications of traditional graded examinations and alternative forms of assessment in different cultures and musical forms are discussed. The development and use of graded examinations for the assessment of Performance Awards in Chinese Music in the ASEAN region is described and the assessment of these awards and the maintenance of standards by assessors from Singapore and Malaysia reported on.

An overview of the history of Graded Examinations

Graded examinations have their roots in forms of apprenticeship and training that are normal in the performing arts. These forms belong to schools of practice and reflect professional requirements and particular traditions. Dance provides examples of this as different traditions and styles such as Russian or Classical French in ballet profoundly influence

both training and performance. People tend to enter into these schools of practice at an early age. This is necessary in order to master the disciplines of an instrument or performance. A proportion of these entrants, progressively attain levels of mastery that enable them to become performers either as professionals or as good amateurs and as 'private' teachers. A smaller number of these will qualify as school teachers and work as peripatetic instrumental tutors or academic music teachers in schools and colleges. All of these but especially teachers in private practice will then pass on the disciplines of their school of practice to another generation of performers. The structure of graded examinations, Diplomas, Licentiateships, Associateships, and Fellowships provide recognition of the progressive mastery implicit in this. This results in the development of a strong sense of belonging to a particular community of practice and a real sense of loyalty to the philosophical traditions that underpin it. This in its turn leads to differences in teaching and assessment and results in differences between examining boards made more on the grounds of culture and values than on the level of technique and expertise required at each grade.

All of the graded examination boards were established well before the transition to competitive examinations such as the College of Preceptors Leaving Certificate described by Wardle (1970)ⁱ or by the School Certificate and Higher School Certificate introduced for general use in schools in 1917. For example, Trinity College of Music dates from 1872, the London College of Music from 1887 and the Associated Board of the Royal Schools of Music was founded and started examining in the UK in 1889. The basic practices, ethos and core activities of these graded examining boards, has remained the same from their inception up until the present day. Changes to repertoire and to the extension of qualifications to the wider public have taken place, first in the United Kingdom and subsequently overseas over the same period. Obvious examples of this

are the introduction of examinations for performers playing electronic instruments and the extension of a western classical musical emphasis to the inclusion of other forms such as Jazz. In the past year the London College of Music (now part of Thames Valley University) has started to offer Graded Awards in Chinese Music that are performed in Chinese and assessed by examiners from the Chinese Music community. Despite such changes, continuity of purpose and its transmission through a community of practice that includes both teachers and examiners is clearly present and recognised by both. It should not be thought that this stability has arisen because of small numbers or exclusivity as the number of examination entries each year, indicate that this is not the case. The Associated Board of the Royal Schools of Music claims that it examines over 570,000 candidates in over 80 countries each year, including approximately 350,000 candidates in the UK. So although no exact figures are available at the time of writing it seems likely that the total of graded examinations in music, speech, drama, and dance from all the UK based graded examination boards is probably in excess of one million a year. Given the numbers of candidates involved and the time that these examinations have been in use it is curious that so little research into them exists.

Graded examinations of performance exist in a range of disciplines where the demonstration of progressive levels of mastery differentiates between the novice and an accomplished performer. Eight grades (plus higher level awards) are used to define increasing mastery in the skills, techniques and knowledge required of a musician. A key characteristic of such examinations is the requirement for a live performance to be given in a 'real world' context.

Progressive mastery emphasises both the continuity of learning and the demonstration of performance at a given standard. Progression results

from requiring the demonstration of skills, techniques, knowledge and understanding. The breadth, depth and relationship between these qualities alters according to the level of mastery that is expected at a given stage in the development of a practitioner. For a graded examination of performance the progressive mastery model results in a structure of requirements that is sequential but not necessarily linear and that incorporates and extends previous learning. Within an examination differentiation is by both content and by outcome. Content is determined by repertoire selected to represent an appropriate level of difficulty for the grade to be examined. Outcome is determined by the demonstration of performance expectations that match or exceed the minimum levels of mastery required. In relation to graded examinations, the result is a series of thresholds for performance and an iterative process that has strong formative assessment characteristics. As a consequence, graded examinations have a different purpose and rationale from that of summative examinations such as General Certificate of Secondary Education (GCSE) and General Certificate of Education (GCE) used in England. This is reflected in the nature of the examination and the way performance is assessed.

For GCSE and GCE examinations the assessment culture has its roots in nineteenth century competitive examinations such as the College of Preceptors Leaving Certificate, Oxford and Cambridge Local Examinations and those for the civil service (Wardle 1970)ⁱⁱ These examinations rank ordered candidates by results, a practice that continues to influence the way in which these examinations have evolved. The assessment culture for Graded Examinations has evolved in entirely different ways. These examinations have their roots in the forms of apprenticeship and training used in the performing arts. These forms belong to schools of practice and reflect professional requirements and particular traditions.

Examples of this may be drawn from dance, where different traditions and styles profoundly influence both training and performance. What is true of dance also applies to the other performing arts. People tend to enter into these schools of practice at an early age. This is necessary in order to master the disciplines of an instrument or performance. A proportion of these entrants, progressively attain levels of mastery that enable them to become practitioners and teachers. These then pass on the disciplines of their school of practice to another generation of performers. The structure of graded examinations, Diplomas, Licentiatehips, Associateships, and Fellowships provide recognition of this progressive mastery. At the higher levels these qualifications have both professional credibility and wide acceptance by those involved in the performing arts. Traditional Graded Examinations of Performance are in effect, a form of qualifying examination for a professional association. This means that the purpose of these examinations has never been competitive selection or rank ordering by results. It also means that these examinations have been conducted by relatively small cadres of examiners; who are themselves usually selected from people who 'belong' to the awarding body by virtue of their own training and qualifications. As a consequence each cadre of examiners tends to be homogeneous and to have shared values, normative understandings and standards of practice. These values and understandings are also shared with the teachers preparing candidates for examination. This is because the teachers also tend to 'belong' to the same awarding body, by virtue of their own training and qualifications as practitioners. The result is a form of collegiate understanding about standards and expectations that is transmitted by training as a practitioner and subsequently reinforced, by entering candidates for the examinations of the same awarding body. In this situation continuity of practice is clearly important, as the maintenance of expectations and standards is effectively vested in the examiners. Consequently, changes to the examiner

population and the introduction of new syllabuses are likely to lead to instability, to questions about the results of assessments and to become a source of unreliability.

In a Graded Examination syllabus such as music, there is a clear progression in the level of demand placed on a candidate. The knowledge, understanding and skills required at each grade, are made explicit through the repertoire selected to be taught and examined. This repertoire is differentiated by required mastery. Required mastery is defined by expert judgement and takes into account a range of factors including the interactions of techniques, skills, knowledge and interpretation, appropriateness of content, levels of expectation and what is to be taught. The result is a selection of pieces that teachers and examiners agree are appropriate for the grade to be examined. Differentiation by required mastery seems to be based on a process of norm based judgements mediated by experience similar to that described as Limen Referencing (Christie and Forrest 1980).ⁱⁱⁱ In this case the process is used to closely define the content available for teaching and examination at each grade. This is then published as a syllabus in which content is explicitly specified and standards and expectations are more implicitly understood by virtue of the content, than specified through aims, objectives, descriptions, or criteria. The purpose of a Graded Examination syllabus is to provide differentiated content and an assessment of the extent to which it has been mastered.

Assessments of performance

Graded examinations are characterised by a single examination of a performance given by a candidate and simultaneously examined by an external examiner. The examination may include supplementary tests of technique, knowledge, and understanding but these are subordinate to the

performance of the repertoire. The only exceptions to this are written tests of theory, taken as separate examinations and the use of supplementary journals in some forms of examination.

Assessments of performance rely on assessor judgement, not least because of the variable nature of performance outcomes. The outcome and meaning of a performance is highly dependent on its context and the interactions that take place between performer and assessor or audience. This means that each performance is a unique occasion, whether given as a second performance of the same work, or when subsequently seen or heard as a recording. For examinations that are wholly or largely performance based, the quality and consistency of assessor interpretation and judgement is of greatest significance in determining the dependability of the examination. Dependability may be defined as meaning that the validity of the content of an examination and the reliability and comparability of its assessment, are at optimal levels in relation to a particular purpose and, to quote (Gipps 1994)^{iv} is:

"The extent to which an assessment would produce the same, or similar score if it was given by two different assessors, or given a second time to the same pupil using the same assessor".

Assessor judgements are arrived at through a process of shared values and knowledge that is tacitly applied throughout the examination process and transmitted in the form of professional understandings. This is the view put forward by Christie and Forrest (1980),^v in relation to the way in which GCE examiners worked, as they formed a series of norm referenced judgements, mediated by past experience. It is probable that the use of written assessment criteria helps to formalise this process but does not change the way judgements are arrived at (Morgan 1996,^{vi} Brown 1996,^{vii} Macintosh & Wilmot 1995^{viii}). Where an examination

requires interpretative or open-ended responses, a greater degree of judgement is required from the assessor. More open-ended or interpretative responses require more interpretation on the part of the assessor. Even where explicit criteria are used, the assumptions, and interpretations made by assessors are rarely accessible to others. In this situation any form of moderation, re-marking or review, that takes into account and relies on the repeatability of assessment judgements, will be based on assumptions and interpretations about the original assessment that are open to question. This emphasises the fact that, unless the tacit assumptions and implicit values that are at work in interpreting criteria, are made explicit, each stage of assessment will add to the possibility of misrepresentation of value.

The influence of the teacher

Similar forms of tacit assumptions and implicit values underpin the contribution of the teacher and its influence on the performance given by the learner during an examination. Whilst the role of the teacher can be described it is fair to say that the extent of teacher influence on the nature of the performance given is a matter of conjecture. Anecdotal evidence from experienced examiners of music, suggests that different teacher abilities, expectations and styles do make a difference to the nature of the performance given. During an examination, the wrong (in the sense of being outside of normal bounds or expectations), interpretations of a score, or a lack of confidence about performance directions may simply represent a faithful repetition of what has been taught and prepared. At one level this observation is simply a matter of common sense, but at another it raises more significant issues. Put crudely the question becomes 'To what extent is it the teacher who determines the final mark that is awarded, rather than the candidate in the examination room?'

This is actually more complex than it appears because the tacit assumptions and implicit values that are involved may (for example) include those drawn from different cultures (e.g. a western examiner working in an Asian context), use of language, non-verbal cues and other affective factors which have the capacity to impact on each of the parties involved in preparing for and participating in the examination process. In passing we should also note the shift in terminology from learner, pupil or student to candidate for examination. This is not unimportant as the examiner also influences the performance given, if only through the atmosphere created, the order in which the component parts of the examination take place and the nature of directions given.

For those not familiar with the way these examinations are conducted the following description may be helpful. Picture a room in a large house or a church building. Near the window, there is a baby grand piano and between this and the door a small table. The door opens and we hear the sound of someone in another room playing a piano. The examiner looks up from the paperwork on the table and smiling, rises to greet the candidate. As the door is closed, the sounds of piano playing fade. Returning to the table the examiner checks the details of this next examination candidate, asking her how the journey had been and putting her at ease. They cross to the piano, and once seated, the candidate gets her music ready and waits, hands in lap, for the examiner to ask her to play the first piece of the music that she has practised many times over the past months.

This brief scene is enacted over and over again with many different instruments and with learners whose abilities, interests and ages differ greatly. What they all have in common is the presence of the teacher who has prepared them for this brief opportunity to perform.

When a music teacher prepares a learner for examination the starting point is usually the choice of examining board. This is frequently pre-selected as many (but not all) teachers owe their allegiance to the examining board or society from which they gained their qualifications. These qualifications (usually in their main instrument) are primarily awarded for performance, and although the theory of music is a prerequisite for higher level awards there is no requirement to recognise ability as a *teacher*. Some higher level awards are made for teaching rather than performance but there is little significant emphasis on pedagogy or the sort of content that informs the majority of teaching qualifications. Custom, practice and repertoire define content and the mode of delivery is primarily through *instruction*. One consequence of this is a tendency to focus on 'learning the pieces' and on matters of control and technique. Although musicality and interpretation is expected it is not unusual for the interpretative aspects of performance to be determined by the teacher, sometimes with unfortunate results for the quality of the performance presented in the examination. The extent of this and the impact of the instructional model and the extent to which it may determine the result for a candidate is unclear, as is the more general influence of the teacher and other factors that have already been referred to.

What emerges from a review of the literature on assessment of musical performance is that there remains a dearth of research in this field. Given the attention paid to links between music and creativity, ability in music as a predictor of other qualities, as a motivator and into functions of music in education this is curious. The reasons why published research skirts this need to engage with the issues of judgements and performance are unclear and go well beyond the scope of this paper. Perhaps the observation by Odham (2001)^{ix} that:

“The formal assessment of musical performance has a long history and practice but, until recently has remained a closed and arcane process.” (p.84)

is part of the explanation. Hallam (2001)^x in a discussion of research into the development of expertise in young musicians, concludes that:

“... research highlights the complexity of the factors that affect progress and the outcomes of learning to play a musical instrument reinforcing the need for multi-dimensional rather than uni-dimensional explanations.” (p. 22)

What is being illustrated is just how complex the business of assessing performance actually is and as a consequence why apparently objective measures of reliability that are frequently published for some forms of assessment should be treated with caution in respect of these. Levels of statistical significance and measures of correlation in regard to assessments of performance are suggestive of a degree of validity that does not exist in reality. Further research is needed to explore these issues and the complex interactions that are part of the examining process, particularly in relation to the meanings attached to reliability and the ways we may choose to assess performance. (William 1996)^{xi} points out:

“... those who do not share the assumptions of the community will not agree about the meanings of the standards.” (p.303)

this is an important reminder that whilst research of the kind referred to previously is necessary for a variety of reasons in the context of graded examinations and assessments of performance; there is no such thing as an ‘absolute’ standard but only a conceptual referent given meaning by those that use it and the community of practice or culture to which they belong.

It is this last statement that forms the basis for the development of new applications of the graded examinations as described in this paper. The evolution of these awards (which are over 150 years old) has taken another step forward with the development of Performance Awards in Chinese Music by one of the first graded examination boards. Taught and assessed in Chinese these awards break new ground by departing from a form dominated by Western styles of music and (more radically) from a way of thinking that is dominated by the imperial and colonial past of which the examinations were originally a part. The development of these new awards also incorporates new assessment and standardisation procedures^{xii} which have enabled standards to be maintained in a cross-cultural setting with assessor training and examinations conducted in Chinese.

Case Study – Performance Awards in Chinese Music

Meetings with Chinese Musicians in Singapore in the early part of 2006 led to proposals to adapt traditional graded examinations in Western Music in order to provide awards for Chinese Instrumentalists and Singers. A key requirement was that these new awards should allow a wide range of repertoire and ensure that both amateur and professional musicians were able to gain recognition of a recognised international standard with assessments in Chinese by Chinese musicians trained as assessors.

Work started in May 2006 with a group of experienced Chinese musicians who were all recognised by the Chinese musical community in Singapore as possessing the necessary status and experience to contribute to the development of these awards. Some of the group had first hand experience of preparing students for graded examinations in Western

music one of the group had trained as a classical Western opera singer in Italy as well as being an accomplished singer in a Chinese music tradition. An initial meeting was spent exploring the practicalities of creating the new awards and in describing the expectations of the participants and possible limitations on the forms of performance and the conduct of assessment. This was followed several weeks later with two days of intensive workshops and discussions facilitated by the authors of this paper. These meetings were mainly conducted in Chinese with translation to and from English as required. The outcome was a clear framework of eight graded awards and progression to the Diplomas offered by the London College of Music. This framework described the initial expectations of mastery for performance at each award and included assessment constructs agreed by the group, rubrics for the conduct of assessments and a requirement for a discussion about the performance with a candidate for awards above Grade 5.

Further work to refine the framework and its contents took place by email and was facilitated by the group convenor in Singapore. This was followed in September 2006 by three days of meetings, workshops and assessment training during which assessments of live performances at various grades (Instrumental and Voice) were conducted by the group working together. Each set of assessments was discussed and reviewed using a form of facilitated consensus moderation in order to establish the standards of mastery and confirm performance expectations for each grade. Another period of consultation by email followed this before a final training session in late November 2006. This was immediately followed by two days of assessments for a group of candidates whose performances in different styles of Chinese music (Instrumental and Vocal) was examined by the newly qualified assessors. These assessments were conducted in Chinese with scores recorded, checked and standardised using a form of group moderation under the direction of Professor John

Howard of Thames Valley University – London College of Music Examinations. Shortly afterwards 69 successful candidates with ages ranging from 14 year to 70+ were presented with their Awards at a presentation in the DBS Auditorium in Singapore.

Members of the original group provided training for new assessors from the Chinese community in Malaysia during January 2007. These sessions were partly facilitated by the authors of this paper but by June 2007 a programme of assessments and assessment training was being provided by the original group and assessors from Malaysia, the standardisation and monitoring of results remains under the control of Thames Valley University – London College of Music Examinations.

A significant factor in the successful development of these awards has been the control exercised by the Chinese musicians and the sense of ownership that this engendered. This has been further enhanced by the reputations and experience of the group both as musicians and teachers as this provided the credibility necessary for acceptance of these awards within the wider Chinese community. At the time of writing the group convenor is involved in discussions with national organisations in China and it is anticipated that these awards will be launched collaboratively in China in the near future. To date in excess of 350 assessments have been carried out in Singapore and Malaysia.

References

- ⁱ Wardle D (1970) English Popular Education 1780 - 1970. Cambridge University Press. p. 95
- ⁱⁱ *ibid*
- ⁱⁱⁱ Christie, T., & Forrest, G. M. (1980) Standards at GCE A-level: 1963 and 1973. London: Macmillan and Schools Council.
- ^{iv} Gipps C (1994) Beyond Testing Towards a theory of educational assessment. The Falmer Press. p. 172-173
- ^v Christie T and Forrest G M, *ibid* pp. 56-66
- ^{vi} Morgan C (1996) - The Teacher as Examiner: the case of mathematics coursework, *Assessment in Education* 3.3 pp. 355-357
- ^{vii} Brown C R et al (1996) The Construct Validity and Context Dependency of Teacher Assessment of Practical Skills in Some Pre-University Level Science Examinations. *Assessment in Education* 3.3 pp. 381,388-389
- ^{viii} Macintosh H G and Wilmot J (1995) A Survey of Current International Trends and Issues in Testing and Assessment. Unpublished paper prepared for the Schools Curriculum and Assessment Authority. pp.12-16
- ^{ix} Odham, G., (2001) Research in the Arts: issues in the development of new and relevant techniques of arts research in music, the arts and arts education. *Music Education Research*, 3-1 2001.
- ^x Hallam, S., (2001) The Development of Expertise in Young Musicians: Strategy use, Knowledge Acquisition and Individual Diversity. *Music Education Research*, 3-1 2001.
- ^{xi} William, D., (1996) Meanings and Consequences in Standard Setting, *Assessment in Education: principles, policy & practice*, 3.3
- ^{xii} The Affirmative Assessment System™ and related software