

New demands, changing values: applications of connoisseurship to the assessment of personal qualities and attributes in diverse cultural and educational settings.

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

Lead author and presenter:

Dr. Jonathan H Robbins

Correspondence to:

The Talent Centre Nottingham Weymouth Dorset DT3 4BH England
+44 (0)1305 814820 robbinsj@talent-centre.com www.talent-centre.com

Second author:

Mr. Carl Emery
Centre for Conflict Learning Education and Research
Manchester M12 4FA

Abstract

Education and employment are changing. New technologies and new working practices, globalisation, and changing markets mean that there is an increasing need for people who are adaptable, willing to learn, enterprising and who are capable of employing a range of transferable skills. Assessment of these and similar attributes requires alternative forms of assessment to those relying on specifications and the production of evidence or on more conventional types of examination. This paper describes the development and application of an approach employing connoisseurship and construct referencing for individual assessments and awards within the Personal, Employability, Achievement, and Reflection for Learning Programme (PEARL) developed with Manchester College of Arts & Technology, England. Amongst the attributes being assessed are willingness to learn, commitment, dependability, self motivation, team work, communication skills, co-operation, drive and self management. This alternative mode of assessment currently focuses on social mastery in diverse cultural and educational settings and is seen as having the capacity to recognise achievement and encourage involvement in learning, particularly amongst marginalised young people. The maintenance of standards and reliability in this alternative form of assessment is discussed and a case study presented, together with a report and evaluation on applications, outcomes, and future development.

Authors Note

The contribution by Vice-Principal Jack Carney and colleagues at Manchester College of Arts & Technology is gratefully acknowledged. Particular thanks are due to Carl Emery of the Centre for Conflict Learning Education and Research, whose work I have drawn on extensively for this paper and whose experience and knowledge has ensured the successful development and delivery of this project. I am also indebted to Dr Nor Idayu

Mahat of the Faculty of Quantitative Sciences, Universiti Utara Malaysia for her contribution to this paper in respect of work on applications of statistical process control.

The Personal, Employability, Achievement, and Reflection for Learning Programme (PEARL).

Personal Employability Achievement and Reflection for Learning awards are offered by the Graded Qualifications Alliance¹ and recognise achievement and ability for employment, and social and personal development. These include (but are not limited to) skills of organisational, social, cultural awareness, problem-solving and creativity. The awards are accessible to learners in schools and colleges, marginalised groups and those who are already in employment. The awards use the progressive mastery model of learning common in Graded Examinations of Music, Dance and Drama. Assessment is of performance and uses a 'show me what you can do and tell me why you did it' approach which encourages assessment for learning and enables a final assessment by an independent assessor. Assessments take into account the learners performance, underpinning knowledge and performance over time. Programmes of study involve learners in discussions, group work tasks, practical activities and interviews. There are two awards at Entry Level and eight Grades covering National Qualification Framework Levels 1 – 3 that applies to England, Wales and Northern Ireland. Learners enter for assessment at the Grade that is appropriate to their level of ability. Awards of a Pass, Pass with Merit or Pass with Distinction are available to those who match or exceed the minimum standards at each Grade. Accredited in 2007 by the Qualifications and Curriculum Authority as a pilot, at the time of writing approximately 2000 participants have either completed or are currently participating in the programme after its launch in Manchester, England.

The Learning and Skills Council (National Employers Skills Survey 2004: key findings)² report that employers note where skills shortages exist in applicants they may be ranked

¹ The Graded Qualifications Alliance is an Awarding Body accredited by the Qualifications and Curriculum Authority (England) that provides awards and qualifications focusing on the mastery of performance. Its Full Members are: The British Theatre Dance Association, Manchester College of Arts & Technology and the University of England.

² <http://readingroom.lsc.gov.uk/Lsc/2006/research/commissioned/nationalemployersskillssurvey2005keyfindings-re-june2006.pdf>

thus; technical and practical skills (47%), communication skills (40%), customer handling skills (36%), team working skills (32%) and problem solving skills (29%).

Top 10 competencies sought by employers based on website search by Edwards (2001)³.

1. Flexibility, adaptability and the capacity to cope with and manage change (88%)
2. Self-motivation and drive (88%)
3. Analytical ability and decision making (75%)
4. Communication and Interpersonal skills (75%)
5. Team working ability and skills (63%)
6. Organisation, planning and prioritisation abilities (50%)
7. Customer focus and service orientation (25%)
8. Ability to innovate (25%)
9. Mental and physical resilience (25%)
10. Leadership ability (25%)⁴

Office for Standards in Education, Children's Services and Skills (Ofsted) research into the relevance of education for the workplace) reported in June 2005⁵ that of 544 20 to 30 year olds questioned regarding how well they perceived that their education had prepared them for working life only 53% said they were well prepared with 45% reporting they were not well prepared. 91% of respondents believed that education should prepare them for working life.

It is clear from the reactions of young people themselves, educators and employers that the current system is struggling to produce young people equipped to enter the world of work as fully prepared as is desired. Currently countries such as India and China can compete on the basis of lower labour costs. But with around 20 million graduates in China and 2

³ Edwards A. M. & Clear, F. (2001) Supporting the Collaborative Learning of Practical Skills with Computermediated Communications Technology, Educational Technology & Society, 4(1)

⁴ <http://www.lsbu.ac.uk/caxton/tutoring-handbook/research/employability.html>

⁵ http://www.ofsted.gov.uk/portal/site/Internet/menuitem.eace3f09a603f6d9c3172a8a08c08a0c/?vgnextoid=c_a102a5f74f3c010VgnVCM1000003507640aRCRD&vgnnextchannel=162d764e0f32b010VgnVCM100000960f430aRCRD

million new graduates each year in India, those countries are increasingly competing not just on cost but on expertise.(Skills in the global economy: HM Treasury et al 2004)⁶.

A good example of the need for highly trained effective workers can be seen in the developments taking place in cities such as Manchester. Here, it is estimated that an additional 58,000 jobs will become available through accelerated growth from 2002 to 2015. (Accelerating the Economic Growth of the North : Manchester Enterprises 2005)⁷. There is clearly an urgent need to ensure that Manchester's employees and young people possess the skills to engage in these opportunities, particularly in the context of the city's aspirations to become a Knowledge Capital.

In the education sector new understanding about when and how we learn, the need for personalised learning and innovation and the building of a national entitlement for learners present a range of new challenges. International debates and developments indicate that development of basic skills and key competencies by the learners in the future will be the overall objective of education (Rye and Torbjørnsen 2004)⁸. In 1996 the United Nations Educational, Scientific and Cultural Organization (UNESCO) International Commission on Education for the Twenty-first Century⁹ argued that education throughout life has four pillars: learning to know, learning to do, learning to live together and learning to be (Delors 1996)¹⁰ This is a view supported by the Organisation for Economic Cooperation and Development (OECD), which recognised a series of key competencies that are essential for the personal and social development of people in modern, complex societies:

⁶ http://hm-treasury.gov.uk/media/2/B/pbr04skills_410.pdf

⁷ <http://www.manchester-enterprises.com/strategy.htm>

⁸ Quoted in Background paper to Workshop 3: Quality Education and Competencies for Life Competencies for Life: Some Implications for Education by Halfdan Farstad National Institute of Technology, UNESCO 47th International Conference on Education Quality education for all young people: Challenges, trends and priorities Geneva, 8 – 11 September 2004
<http://www.ibe.unesco.org/International/ICE47/english/Organisation/Workshops/Background%20at-3%20ENG.pdf>

⁹ <http://www.unesco.org/delors/>

¹⁰ *ibid*

Each of these key competencies implies the mobilisation of knowledge, cognitive and practical skills, and social and behavioural components including attitudes, emotions, and values and motivations. (OECD 2003)¹¹

Indeed a study conducted by Eurydice in 2002¹² highlighted two clear motives for emphasis to be placed on key competencies:

“Firstly, quality in education, stemming from variation in educational attainment from school to school. Secondly, internationalisation, rapid scientific and technological progress, and increasingly complex career paths”

Key competencies include attitudes, knowledge and skills that are not specific to particular subjects. Hence the learning activities connected to their development must be adapted to the various subjects and levels. For example, in each subject, the student is expected to build not only subject competence, but general learning competency by learning to set targets, to plan the learning activities, to carry out the learning activities and to assess his/her academic performance, achievements and products. Key competencies have been identified as Managing Feelings, Social Skills, Problem Solving, Motivation, Change Management and Self Awareness.

It is clear that personal, social and emotional development is being recognised as a key area for all learners to explore, engage with and build on during their schooling and through into employment as part of the life long learning process. Learners require a range of key transferable skills that will not only prepare them for education, work and life but that are also measurable, valued and relevant.

In education these developments can be seen in the work of the UK’s Department for Education and Skills (DfES) and the Qualifications and Curriculum Authority (QCA), which have both developed national frameworks around social and personal development. The DfES through the Social and Emotional Aspects of Learning (SEAL) programme at the Primary level and the Social, Emotional and Behavioural Skills (SEBS) programme at the

¹¹ Key Competencies for a Successful Life and a Well-Functioning Society
http://www.oecd.org/document/50/0,2340,en_2649_201185_11446898_1_1_1_1,00.html

¹² <http://www.eurydice.org/portal/page/portal/Eurydice/showPresentation?pubid=032EN>

Secondary level. The QCA through Personal, Learning and Thinking Skills (PLTS), part of the 14-19 education strategy.

These frameworks are intended as a catalyst for schools and colleges to develop their own work around social and personal competencies. These are not the only programmes. The teaching and training sector now presents a wide range of schemes to deliver and build on learners' social and personal development. These include programmes such as the Royal Society for the Arts: Opening Minds Programme, The ASDAN: Certificate of Personal Effectiveness, the Duke of Edinburgh Award and models from the U.S. such as Promoting Alternative Thinking Strategies and Second Step.

These goals can be seen in other practical developments such as the introduction of the specialised Diploma and the role of personal development in the UK Government 14-19 White Paper. The specialised Diploma, developed in partnership between the DfES, QCA and Skills for Business Network, is a combination of practical skills, theoretical knowledge and technical ability. One of the stated aims of the Diploma is to:

“provide credible, industry verified applied learning, linked to interdependent general learning, with real opportunities to practice skills” QCA 2006

The structure of the Diploma will ensure that learners undertake generic learning covering functional skills (Maths, English, IT) as well as personal, employability and thinking skills including team working, enquiry, self management and communication skills. In the UK a study commissioned by the Department for Education and Skills highlighted a number of specific benefits that teaching these skills can bring. In summary, these were identified as:

- greater educational and work success;
- improvements in behaviour;
- increased inclusion;
- improved learning;
- greater social cohesion, increase in social capital;
- improvements to mental health.

The GQA PEARL programme is a pioneering learning framework for assessing, delivering and measuring employability and the personal competencies or development of the learner. It is not a teaching and learning syllabus with curriculum content but rather a teaching and learning framework which also acts as an assessment model. In essence it is a tool for modelling and building employability and personal development skills. As previously stated there are already a wide range of teaching and learning materials in this field, including fully developed curricula.

There are a wide and varied number of approaches to delivering social, emotional, employability and personal development programmes. However, it is clear that if one is to endorse and develop the aims of these programmes, assessment and recording is vital on a number of levels including: learner progression and awareness, subject understanding, learner self-esteem, school / organisational standards, teaching and learning and in establishing and promoting the validity and academic rigorousness of the subject field. As earlier commentators have noted (e.g. Weare and Gray 2003) we are still a long way from having long-term valid UK data for the effectiveness of social, emotional and personal development programmes, particularly with regards to academic performance.

A number of specific personal, social and emotional competency assessment tools exist including Elizabeth Morris's Whole School, Class and Individual Emotional Literacy Indicators, NFER Nelson's Emotional Literacy – Assessment and Intervention Programme, Schutte's Emotional Intelligence Scale, Petrides, Furnham and Frederickson's Trait Emotional Intelligence Questionnaire and the Mayer-Salovey-Caruso Emotional Intelligence Tests (MSCEIT) Emotional Intelligence Test. However, it should be noted that these assessment tools do not necessarily cover the full range of the programmes being used in the UK. They are also based on very different definitions and perspectives of what personal, social and emotional competency means and generally have a specific focus i.e. assessment for pupils, assessment for parents, assessment for specific ages.

These programmes are based on a number of assessment techniques designed to measure personal, social and emotional competencies. Clearly, these techniques employ a variety of methods reflecting the range of definitions and perspectives on the subject. Common methods include self-reporting (pupils reflecting on their own ability to meet a number of personality traits); observer reports/360 degree assessment (asking a range of others to make a judgement on pupils behaviour, e.g. can verbally express a range of feelings); and finally ability tests (directly testing the respondent's skill in this area and then subjecting the response to a cross-referenced judgement).

When assessing or measuring personal, social and emotional competence it is generally acknowledged that there are two overarching perspectives. Some commentators pursue the trait (personality) definition of social and emotional literacy. This perspective sees skills such as managing feelings and social skills as inherent characteristics of personality. Thus, they should not be judged by others and they cannot be clearly defined and categorised. Consequently, assessing these skills involves self-reporting and learner perceptions of themselves. This view is supported by Petrides et al.¹³ who have based their Trait Emotional Intelligence Questionnaire around this.

The alternative view, pursued by Mayer, Salovey and Caruso¹⁴, is the ability model, namely that such skills as managing feelings, motivation and empathy can be learned and measured. This perspective promotes a cognitive-based theory which considers emotional intelligence as a measurable construct based on performance indicators.

The GQA PEARL Programme is based upon the ability model because:

- research shows that skills such as an ability to manage one's feelings, an ability to manage one's time and changing attitudes are not inherent to personality but can be developed through learning. In addition, this is the perspective, through modeling and

¹³ c.f. [http://www.ioe.ac.uk/schools/phd/kpetrides/Reprints/PAID%20\(2004\)%20-%20T_EI.pdf](http://www.ioe.ac.uk/schools/phd/kpetrides/Reprints/PAID%20(2004)%20-%20T_EI.pdf) and other similar articles

¹⁴ See their MSCEIT: Emotional Intelligence Test for an example of this.

teaching; that the DfES are promoting through SEAL and SEBS, and the QCA through PLTS.

- employers are looking for these key skills and abilities, and it is these that will be measured in job applications and on-the-job performance management, in other words they are the key 'transferable' skills need for the 21C,
- learners have a right to know what level they are at in these key skills and abilities. In any other field, if skills are valued then they are measured and given a 'grade', why not in PEARL?
- the PEARL skills and abilities, although based on performance assessment, are actually process skills integral to self-assessment. For any learner to undertake the self-assessment process they would need to perform the PEARL skills.

An outline of the GQA PEARL assessment programme

It should be noted that the GQA PEARL assessment programme is the only performance based ability model we are aware of in development within the UK. It differs from current provision in that:

- it is performance based
- it is accessible by ability not age
- it brings together the full range of personal and employability skills
- it employs the ability model of measurement
- learners can progress from entry level through to level 3
- it provides a set on step off framework
- the standards are set by the community of practitioners in the field

One of the great benefits of this model is that it:

"...evaluate(s) participants' responses according to a criterion of correctness rather than relying on a self judgement such as "I am emotionally intelligent (True-False). This is

important because actual skill is very different (i.e., uncorrelated or minimally correlated) with self-judged skills in this area" (Mayer 2004)¹⁵

A recent study by Oxford University, *Assessing Emotional and Social Competence in Primary School and Early Years Settings* (2003), undertook a systematic search and critical appraisal of measures of emotional competence that might be suitable for use in preschool and primary school settings. The reports summary identifies a number of key findings.

Firstly, it was found that effective assessment required skilled observers, but that not all teachers want to develop these skills.

Secondly, the report identifies three specific approaches to the assessment of social and emotional competence. These are screening, profiling and monitoring. Screening is recognised as being necessary to identify children who may require specific support in this area. In some instances this takes the form of baseline assessments undertaken by all pupils at the start of a programme.

Profiling involves a variety of observers considering from a range of perspectives what could be done to improve a pupil's emotional and social competence as well as identifying where they have got to so far. Finally, monitoring is used to detect change and identify improvements.

The GQA PEARL model compliments all three of these areas. It empowers skilled observers to quickly and efficiently screen and profile learners against standardised criteria. The system allows for learners to be given reliable feedback and scores on a variety of individual constructs, enabling a focus to be placed on areas that need to be worked on. The results of performance assessment can also be assessed at an organisational level, e.g. school, college or business, to identify key trends and areas for training and development across the board.

¹⁵ "Emotional Intelligence: Theory, Findings, and Implications" John D. Mayer, Peter Salovey, David R. Caruso *Psychological Inquiry*, 2004, Vol. 15, No. 3, Pages 197-215

The monitoring element is supported by periodic assessments of performance and the progressive mastery inherent in the programme, which allows learners to attain a range of grades from Entry Level through to Level 3 of the National Qualifications Framework.

The programme can be used as either a stand alone assessment tool or to give a validation and accreditation framework to other personal, social and emotional literacy programmes such as SEBS, PLTS or Opening Minds.

The programme consists of:

- a) An assessment framework¹⁶
- b) A programme syllabus
- c) A community of practitioners

The assessment framework, including suggested content and examples, is based around the assessment of a set of core skills and abilities per unit. The units are:

Social and Personal Development:

Self awareness, Social skills, Managing Feelings, Empathy, Motivation

Organisational and Creative Development:

Investigation, Time and Resource Management, Planning and Prioritising, Creative Problem Solving, Managing Change

Cultural Development:

Diversity & Inclusion, Stereotypes & Prejudice

For each unit the skill sets or assessment constructs are outlined in graded form, starting at grade B (entry level) through to grade 8 (level 3). These skills and abilities can be assessed

¹⁶ The Affirmative Assessment System™

on a cross-curricular basis or incorporated into a workplace training, management or supervision programme. The emphasis is placed on skills being developed and demonstrated through group work, team challenges, tutor, learner and peer feedback, group discussion, activities, creativity and project work in a variety of settings (classroom, workplace, social setting, home).

The assessment framework is based on a series of 'show and tell' snapshots, perhaps as many as twenty, but centred upon three clear formal assessments where learners perform and are assessed against the criteria set out by the community practicing in this field. In this case, it is the teacher or tutor who has had assessor training that will undertake the formal assessments.

An example of this type of system in action would be a highly respected chef or footballer, acknowledged by their peers as being competent and able, being invited to watch a number of learners in their field of expertise (cooking or football) demonstrate their abilities. These experts through their knowledge and experience know what a competent person in their field looks like, indeed many would say they can "smell it!"

With this knowledge, the practitioner or connoisseur can very quickly make a judgement as to how able the learner is and place this judgement within a set of scoring criteria and grades. By gathering together the views of a range of connoisseurs, a range of examples are set out, thus giving guidance as to what one would expect to see from a learner at each grade.

Consider these:

- Paul McCartney knows a good song when he hears it.
- David Hockney knows a good picture when he sees it.
- Linford Christie knows what a good athlete looks like.
- David Beckham knows what a good footballer looks like.

In the same way a good teacher, tutor, employer or manager knows from regular contact with their learners or staff who is highly motivated, who is organized, can build good relationships, is creative or shows empathy to others.

Assessment only takes place when the learner is ready and is essentially a tool for benchmarking the 'distance travelled' and recognising the achievement of the learner. The assessment process is quick, efficient and clear, based on simple, expert judgements. Clear qualitative data can also be plotted and comparisons made across different parts of the school or organisation, e.g. the ability to identify that one group of learners have far stronger listening skills than their peers or one section of an organisation has very developed skills of empathy.

Responsibility for the assessment judgements is placed upon the teacher, youth worker or manager, as they have the regular relationship with the learner and the close contact needed to build up an understanding of the skills and qualities demonstrated.

In Manchester, a local initiative involving schools, colleges, Higher Education Institutions and employers in the service industries is being used to pilot and evaluate the programme and its assessment. These different sectors are all working together, recognising and validating the performance assessment system for learners and staff. The attraction for key employers such as hotels, theatres and football clubs is the ability to both develop and measure employability skills. At the time of writing, (2007) there are close to 2000 learners taking part, or about to do so from the following organisations:

Manchester College of Arts and Technology,
Manchester Airport,
Manchester South District 14-19 collegiate,
Manchester East District 14 -19 collegiate,
Royal Bank of Scotland,
North Manchester High School for Boys,

The Manchester Standard Group (key employers including City Council, Manchester City Football Club, Lowry Hotels),
New East Manchester Education Team,
Exeter University,
Manchester Adult Education Service,
Cardboard Citizens Project,
East Manchester Complimentary Education Group.

The PEARL scheme is:

- developing good practice in 21st Century Learning, Brain-based learning, Critical and Creative thinking, Accelerated Learning and Assessment for Learning thereby providing a strong basis for an effective Independent Learning assessment framework applicable across all age ranges;
- setting out the first national performance based assessment system for measuring and validating social, emotional competence and employability skills;
- providing a valuable, credible and practical award to support learners' personal development, employability opportunities and their understanding of their own and others' needs;
- adding value to the wide range of social, personal, thinking and employability skills currently being identified, explored and delivered both within and outside of the curriculum;
- progressing a teaching and learning and curriculum framework based on a series of measurable key competencies and skills that place practical demonstration at the heart of learning and employability;

- providing a realistic and valuable award for schools, colleges and workplaces to measure the success of their social, emotional and skills-based programmes, needs provision and skills gaps across the sectors and institutions;
- placing increased importance and value on behaviour, attendance and the understanding of ones own learning through the assessment process;
- promoting and rewarding the application, values and importance of personal and social responsibility, employability skills and cultural understanding;

The maintenance of standards and reliability in performance based assessment

Current institutional adherence to forms of examination, assessment, and reliability owe more to nineteenth civil service entry requirements (Wardle, 1970) and twentieth century concerns with psychometric testing and measures of reliability than to current understandings and practice. This concern with a narrow understanding of reliability in examinations and assessment has been at the cost of focusing on questions of authenticity, validity, and meaning. A great deal of assessment research literature centres on improving the methodologies of well-established assessment techniques, rather than questioning either the validity of the approach itself or its suitability for the purpose in question (Broadfoot & Black, 2004, p.10).

The opportunity to 'show and tell' as a performance and to create authentic learning situations which are complex, involve interactions within an authentic social milieu and which allow for novelties to arise and be assessed creates a demand for new conceptions of reliability and for more appropriate ways to indicate the extent that results and standards are meaningful for all users. As Griffin (1998) argues of outcomes based education, the emphasis should be on how the learner is able to show they have changed as a result of the learning experience rather than focusing on inauthentic tasks that they perform to demonstrate their learning.

Making assessments of performance is extremely complex. As a consequence the apparently objective measures of reliability that are frequently used and published in performance-based assessments should be treated with caution. Levels of statistical significance and measures of correlation in regard to assessments of performance are suggestive of a degree of accuracy that may not exist in reality. To paraphrase a comment by Gipps (1994) measures of reliability should be set in the explicit context of what is or is not being valued and on the basis of what evidence or prejudice is known to be present.

Until recently much of the work on reliability has focussed on issues related to rank ordering and comparability, mainly drawn from a psychometric tradition and employing techniques based on the analysis of variance being widely used and reported on. However, the assumptions underlying these techniques and the ways in which they are applied has been questioned, particularly with regard to assessments using forms of construct referencing (William, 1997).

By their very nature, construct-referenced assessment outcomes are discrete rather than continuous. Outcomes may be reported as pass-fail, or on some scale of grades, and, as the literature on criterion-referenced assessment shows very clearly, indices of reliability that are appropriate for continuous measures can be highly misleading when applied to discrete scales.

Although we can say what reliability ought to mean, we cannot say what it is until we can say with some certainty what it is not. If this is to be achieved, then the methods and measures used must be sufficiently rigorous to stand comparison with the concept of reliability that is predominately rooted in a psychometric and quantitative paradigm. This is a fundamental reason why it is necessary to identify and analyse sources of unreliability in the assessment of performance. By itself, this is not enough, as it is also necessary to put into place objective measures of reliability that are both appropriate to the nature and purposes of the assessment and to use measures that enable comparisons to be made both within and outside of the predominately personal and qualitative paradigm. Two conditions are necessary for this. First, what is required are measures that are tried and

tested in such a way that their validity is not a cause for concern and that are also capable of being applied to the assessment of performance. Second, the nature of this application means that some form of construct referencing will be in general use and consequently the relevance, quality and repeatability of judgements, must be the focus and not some form of psychometric testing.

the issue in relation to assessments of performance, is not whether one way of quantifying reliability is better than another one, but rather of selecting a way of stating to what extent a particular form of assessment is appropriate to its purpose and can be relied upon. To paraphrase the comment of Wiliam (1997 July) about raters, 'to put it crudely it doesn't matter how reliability is calculated, only that what it tells us is relevant'.

Case Study

The British Theatre Dance Association

Founded in 1972 the British Theatre Dance Association, Leicester based BTDA offers training programmes and examinations in Theatre Dance throughout the UK and overseas. It offers a carefully structured series of awards to suit different ages in the dance disciplines Classical Ballet, Modern, Lyrical, Tap and Greek. BTDA is approved by QCA as an Awarding Body and assesses more than 60,000 students every year.

Each examination is comprised of three components (Barré, Centre and Dance) and is assessed by a single examiner on one occasion. Candidates prepare for examination at the Grade which represents the level of mastery appropriate to them. The knowledge, understanding, skills and performance abilities at each Grade subsume those that have gone before and prepare the candidate for the next stage in their development. Up to four candidates are assessed simultaneously using assessment categories (constructs) and criteria, scores from these are aggregated to produce a final score expressed as a percentage. British Theatre Dance Association Graded Examinations in Theatre Dance were developed from similar graded examinations offered by the British Theatre Dance Association (BTDA) over a thirty-year period prior to the introduction of the new syllabus. This syllabus and the examination have been formed around the core practice and the

programme of study that existed but with the application of a new marking scheme and method of assessment¹⁷. The development of this and the transition to the new assessment system took place over a period of twelve months, involved all thirty examiners and represented a fundamental change, especially in the way examiners worked. The overall application and functions of the assessment system, marking scheme and standardisation processes were designed from the start as an integrated and coherent system intended to provide the management information necessary to manage for optimal dependability (validity and reliability) in assessment. Examinations take place by arrangement and in response to request by teachers and are organised as sessions in which the performance of candidates in any of the disciplines or grades may be examined. A typically session comprises eight to ten examinations that take place on a single day in an approved centre. Up to four candidates entered for the same grade and in the same discipline are assessed simultaneously by a single examiner.

The process of standardising results and the way that it is applied both to assessment and to maintaining the stability of judgements in the context of these examinations is described below.

The purpose of standardisation is to ensure both the stability of examiner judgements and the comparability and consistency of final results, by 'bringing assessments into line' Daugherty (1997). The method described below is a form of standardisation by inspection and analogous to moderation by inspection as described by Gipps (1994). Standardisation by Inspection, focuses on the key issue of the repeatability of examiner judgement, as it is a lack of consistency in judgement that is likely to be the greatest source of unreliability (Spencer 1981). Inspection is carried out by reviewing the detailed information available from six sources:

1. the final score for the candidate and the Assessment Audit Trail that supports it;
2. statistical information about examiner performance over time and in relation to all other examiners;

¹⁷ The Affirmative Assessment System™

3. statistical information about examiner performance in the current examination session;
4. statistical information about the performance of an examination centre over time and in relation to other centres
5. information from a Reference Set comprised of video-taped examinations and 'Referent Scores' for each grade of examination;
6. qualitative information derived from examiners, teachers, discipline professionals and others involved in the community of practice associated with the examination.

Standardisation by inspection is carried out by reviewing the pattern and repeatability of examiner judgements as seen in the:

1. mean of the final scores awarded by an examiner during an examination session at a centre (typically thirty to forty final scores), to the representative mean scores of the examiner over time (the Rater Referent);
2. first referent set in relation to the mean of the mean scores of the universe of examiners (the Examination Referent);
3. mean of the final scores awarded by an examiner during an examination session at a centre, to the mean scores of the same centre over time (the centre referent).

If all these reviews indicate that the pattern and repeatability of examiner judgements for that session are within bounds, the results are accepted as dependable and declared to the candidate (who is also given a copy of the candidate profile generated by the assessment system).

If any of the reviews indicate that the pattern and repeatability of the mean final scores awarded by the examiner for that session are out of bounds the reasons are investigated for:

1. administrative error (by individually checking the examination data records);

2. a marked change in the pattern of mean scores by either the examiner or in the centre records.

If there are no administrative errors and a marked change in the pattern of mean scores, either of the examiner or of centre is confirmed, then the nature of the change is investigated and a correction applied to bring them within bounds if this is judged to be appropriate. If random error or special circumstances are identified, the final scores are reviewed by an external moderator using all the information available from the Assessment Audit Trail, examinations results database and the video archive that is maintained for all examiners, disciplines or subjects and grades. Both the examiner and the centre are contacted and asked to identify any special circumstances that may have contributed to an anomalous result.

This method applies techniques drawn from quality assurance procedures in aircrew training and engineering. It combines them with procedures for consensus moderation and statistical monitoring of examiner performance that are widely used by the GCE and GCSE awarding bodies. The result appears to be at least as rigorous and robust as other methods that are in general use, not least because it enables the awarding body to manage the dependability of an examination, by monitoring and controlling assessment judgements at all stages of the examination process. This is because when using the system, the focus of standardisation is not simply on ensuring comparability of results, but on maintaining the dependability of the examination. This requires both the validity of the examination and the reliability of the assessment process to be managed in order to maintain them at optimal levels.

Standardisation by inspection makes this possible because of the data generated as each examination is completed, as this accumulates the performance of each part of the examination may be analysed and compared. Results that appear anomalous may then be investigated and remedial action taken. Similar procedures may be applied to the performance of examiners, for example if there are differences in interpretation between examiners or if queries are raised about the examination components or assessment

categories. In each case, the purpose of doing this is to optimise the dependability of the examination as a whole and to feed this information back to examiners in order that their judgements remain stable and within bounds.

Standardisation by inspection is underpinned by training at regular intervals to ensure that the marking scheme is being used as intended, that the administrative requirements of the examination are fully understood and that the required referents are appropriate to the purposes of the examination and the needs of wider community of teachers and students.

The process is a form of continuous process improvement and may be summarised as a six step cycle:

1. Monitor the assessment and standardization processes to check that they are generating results that remain within the bounds set for stability.
2. If not investigate and identify those that are (i) variations due to external and uncontrollable influences on the system and those that are, (ii) variations caused by sporadically occurring factors that are 'special causes' of instability in either examination results or system performance.
3. Set priorities for investigation (investigations are to find ways of changing processes, by permanently removing or minimising special causes of variation).
4. Initiate, monitor and demonstrate the effect of process adjustments.
5. See that information about factors that affect the process together with knowledge of agreed changes are immediately available to examiners and others involved in standardisation and administration.
6. Repeat the cycle.

Applications of statistical process control applied to assessments of performance in graded examinations

A form of statistical process control incorporating the principles described above has been applied to assessments of performance in graded examinations as part of an investigation into ways of:

1. presenting information in graphical forms for the purposes of standardisation and quality assurance of expert judgements;
2. monitoring and auditing standards of teaching and assessment at local, regional and national levels.

The information generated by these applications (using software developed for awarding body processes) provides management information that is readily accessible as well as data that may be used for the analysis of scores by candidate, cohort and rater, together with procedures for verification, tests of inter-rater reliability, comparability of assessments and the evaluation of teaching and/or learning strategies. This data may also be analysed using (for example) Multilevel and Generalized Linear Models.

The use of Statistical process control (SPC) performance in graded examinations (as part of this ongoing investigation) is outlined briefly below.

Statistical process control (SPC) is a simple, yet powerful, collection of tools for graphically analysing process data that was invented in the late 1920's by Walter Shewhart in order to monitor and improve processes. Originally intended for use in a manufacturing environment it was subsequently extended by W. Edward Deming to improvement in all areas of an organization. Outside of education, it is well known and widely used, especially in engineering although its applications are by no means limited to that.

Statistical process control uses statistical analysis of individual process measurements to categorise performance variations in one of four ways: common cause, special cause, compensation and structural. Common cause variations are the result of everyday, uncontrollable influences, special cause variations are sporadically occurring factors that send performance outside the range of common cause variation, compensation is a cause

of variation arising from attempts at control and structural variations occur systematically because of cycles or trends. The emphasis of statistical process control is on the reduction or elimination of special cause factors and inadequately controlled compensations, so it provides a means of more accurately measuring performance against flexible standards and by defining common cause variation and the limits of this, establishes acceptable performance ranges. Statistical process control makes extensive use of the graphical representation of quantitative information. Typically, this information is provided as one or more process control charts used to plot a function of process measurements against time. Points that are plotted on the graph are compared to a pair of control limits in order that the process may be both monitored and improved, control charts represent a compromise between the risks of not detecting real changes and of false alarms and for that reason the choice control limits needs careful consideration.

A typical control chart (Figure 1.) is a graphical display of quality characteristics that have been measured from a sample (Y-axis) versus the sample over time (X-axis). It contains three main horizontal lines: a centre, a lower control limit (LCL) and an upper control limit (UCL). A centre line is located in between the control limits where it represents the average value of the quality characteristics corresponding to the in-control state. The upper control limit (UCL) and the lower control limit (LCL) are chosen so that if the process is in-control, then nearly all the points will lie within these two lines. This means the process is in-control and no action for adjusting the process is necessary. Using this chart it can be seen that a point that lies outside of the control limits is interpreted as evidence that the process is out of control. So further investigation is required to find possible causes and some action taken to adjust the process in the light of the information provided.

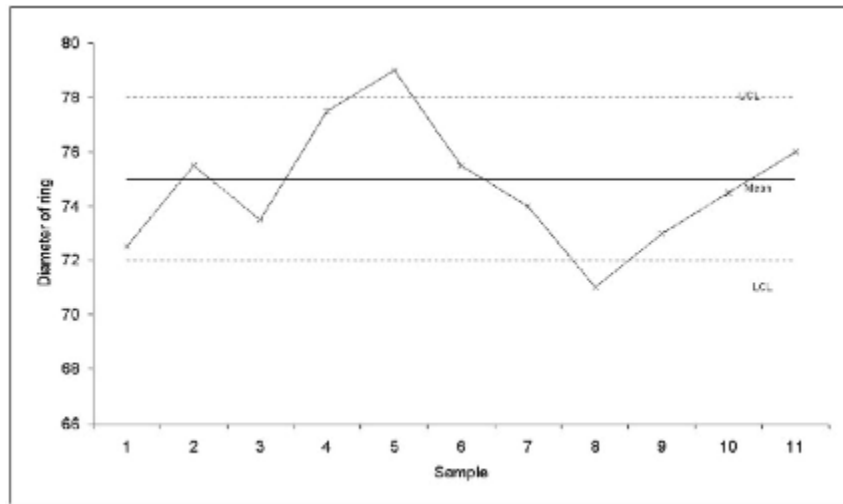


Figure 1: An example of control chart with a centre line and two control limit lines.

The standard Shewhart control charts are widely applicable in product-oriented frameworks because the definition of quality or performance is well defined (Montgomery, 1990). Expanding them into non-manufacturing systems is also possible, however it needs extra caution because most non-manufacturing operations do not have a natural measurement system that allows an analyst to easily define performance parameters, in addition the observability of process in a non-manufacturing setting may be fairly low.

The idea of using these control charts to measure performance of non-manufacturing based is initially comes from the process of conducting hypothesis testing for comparing performance between institutions. In investigating dance centres, given different centres that have conducted the same dance examination, we can test whether they achieve the same performance using a selected indicator by conducting a hypothesis test

H₀ : All centres have the same performance

H₁ : At least one of the centres has different performance.

If the performance indicator comes from a continuum, then the simple analysis of variance (ANOVA) can be conducted. However, analysts need to carry out further analyses to identify which centres are different if H_0 is rejected. Alternative strategies using different forms of control charts may also be employed.

The use of control charts for monitoring performance of dance centres has several advantages. It offers an easy and fundamental way to view data so that any shift in performance can be detected continuously. A system of global monitoring enables performance comparisons to be made between centres. By using this system each centre is able to understand their position in a national table of dance schools and centres (based on rank order) and whether they are performing as well as the average performance of other centres. Meanwhile the local monitoring system provides some information about the centre itself. An awarding body, committees in an examination board, schools or centres may use this information to identify and to investigate the causes of problems, and take action to rectify them.

Evaluating results on an annual, quarterly or monthly basis is possible with some adjustments to the computation of charts and estimators. It is worth remembering that the role of control charts is to display or warn of any possible causes of problems. They do not provide suggestions on how the problem should be handled. Pursuing further investigations to identify the causes of problem are needed, this often needs more qualitative investigation methods as it is not always necessary (or desirable) to use sophisticated mathematical tools.

Research into the use of statistical process control (focusing on evaluating the performance of dance centres in the UK) by Dr Nor Idayu Mahat of the Faculty of Quantitative Sciences, Universiti Utara Malaysia, and the author will be reported on in the near future.

References

Broadfoot, P. & Black, P. (2004) The First Ten Years of Assessment in Education, *Assessment in Education*, 11 (1), pp. 7-27.

Daugherty, R., (1997) Consistency in teachers' assessments: defining the problem, finding the answers. *British Journal of Curriculum and Assessment* 8.1

Delors 1996

Learning: The Treasure Within, Report to UNESCO of the International Commission on Education for the Twenty-First Century, UNESCO, Paris 1996

Edwards A. M. & Clear, F. (2001) Supporting the Collaborative Learning of Practical Skills with Computermediated Communications Technology, *Educational Technology & Society*, 4(1)

Eurydice 2002

Key Competencies: A Developing Concept in General Compulsory Education, Brussels 2002

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

Griffin, P., (1998) Outcomes & Profiles: Changes in Teachers' Assessment Practices, *Curriculum Perspectives*, Vol 18, No. 1, 1998 pp 9-20

HM Treasury et al (2004)

http://hm-treasury.gov.uk/media/2/B/pbr04skills_410.pdf

Mayer (2004)

"Emotional Intelligence: Theory, Findings, and Implications" John D. Mayer, Peter Salovey, David R. Caruso *Psychological Inquiry*, 2004, Vol. 15, No. 3, Pages 197-215

Montgomery, D.C. (1990) *Introduction to Statistical Quality Control*, 2nd edition. (New York: John Wiley & Sons).

OECD (2003) *Key Competencies for a Successful Life and a Well-Functioning Society*

http://www.oecd.org/document/50/0,2340,en_2649_201185_11446898_1_1_1_1,00.html

Petrides et al c.f.

[http://www.ioe.ac.uk/schools/phd/kpetrides/Reprints/PAID%20\(2004\)%20-%20T_EI.pdf](http://www.ioe.ac.uk/schools/phd/kpetrides/Reprints/PAID%20(2004)%20-%20T_EI.pdf)

and other similar articles

Rye and Torbjørnsen (2004) Quoted in Background paper to Workshop 3: Quality Education and Competencies for Life: Some Implications for Education by Halfdan Farstad National Institute of Technology, UNESCO 47th International Conference on Education Quality education for all young people: Challenges, trends and priorities Geneva, 8 – 11 September 2004

Stewart-Brown S, Edmunds L.(2003). Assessing Emotional and Social Competence in Early Years and Primary School Settings. *Perspectives in Education*;21:17-41

Wardle, D. (1970) *English popular education 1980-1970*. Cambridge: Cambridge University Press.

Weare and Gray (2003) *What Works in Developing Children's Emotional and Social Competence and Wellbeing*, The Health Education Unit Research and Graduate School of Education, University of Southampton 2003.

William, D. (1997) Construct-referenced assessment of authentic tasks: alternatives to norms and criteria. <http://www.kcl.ac.uk/depsta/education/hpages/EARLI97.pdf>

William, D., (1997, July) How to do things with assessments: illocutionary speech acts and communities of practice. Paper presented at 21st Conference of the International Group for the Psychology of Mathematics Education Discussion Group on Assessing Open-Ended Work in Mathematics held at Lahti, Finland. London, UK: King's College London School of Education.