

The Moderation of Liberal Studies School Based Assessment Scores: How to Ensure Fairness and Reliability?

Tao Wang YU

E-mail: yu3897@yahoo.com.hk

Abstract

As part of the ongoing education reform initiated by the Hong Kong Education Bureau (EDB), Liberal Studies will become a compulsory subject in senior secondary schools in 2009. It will be one of the core subjects all students must take, besides English, Chinese and Mathematics. Some of the objectives of the subject include cultivating critical thinking, developing generic skills, life-long learning and encouraging students to see things in multi-perspectives. However, aligning the learning goals and assessment will not be an easy task. The traditional forms of assessment may no longer be adequate to assess our students. According to the officials from Education Bureau, the introduction of School Based Assessment (SBA) was deemed more desirable than assessing students by one written examination. But some students are concern about the fairness and reliability of teacher assessments. The issue of moderating SBA project work needs to be addressed. Would teachers be more lenient in judging their students' work? Do teachers have bias? Should the grades of project be based on exam results? How should moderation be done to ensure fairness? The aim of the present article is to discuss the moderation of school based assessment scores in the new senior secondary Liberal Studies subject and the adjustments schools and the authorities might have to make.

Keywords: School Based Assessment, Moderation, Liberal Studies

1. Introduction

According to the EDB, the aims of Liberal Studies are to broaden students' knowledge base and perspectives, and enhance their social awareness through the study of a range of issues. The curriculum is designed to allow students to make connections across different fields of knowledge and to broaden their horizons. The learning experiences provided will foster students' capacity for life-long learning, so that they can face new challenges confidently. The framework of curriculum and assessment takes into account of cross-disciplinary studies, pertaining in particular to critical thinking, life education and values education relevant to the Hong Kong context. To put these ideas into practice will require modifications to the teaching strategies, assessment methods and school culture.

2. The Challenge

Under the new secondary school academic structure, all senior secondary students will study Liberal Studies as a core subject from Form 4 to Form 6, and then sit for a public examination. The exam result may determine whether the student can get a place in university. Some people query that how can the assessment be designed to align with the goals of the subject. For instance, how can we measure a student's ability to think independently? How can we assess generic skills?

During the consultation stage of the new curriculum, there were arguments on whether Liberal Studies students should sit for examinations. Some educators argued that the public exam will destroy the "spirit" of the subject. Teaching-to-the-test and rote-learning will again dominate teaching and learning. Much time will be allocated on preparing exams rather than developing thinking skills and life-long learning attitude etc. As pointed out by Madaus that (as cited in Weeden, Winter and Broadfoot, 2002), when test results are the sole or even a partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than a useful but fallible indicator of achievement. In fact, students should be learning concepts, process, or ideas that lie behind it rather than just test items (English, 2000). On the other hand, some argued that if there is no examination, student will not take the subject seriously. Students only study what is on the test, and may not even try to work hard in this subject. After many consultations, in March 2007, the new curriculum and assessment method of the Liberal Studies is finalized. The following table shows the design of Liberal Studies assessment (EDB, 2007):

Insert Table 1 Here

Cheung and Hui (2006) point out that, in many cases, public examinations assess learning outcomes (not always related to curriculum aims) rather than the learning process and change. Therefore, much effort is spent on training examination techniques instead of the pursuit of the goals of the subject (values and attitudes in the case of Liberal Studies). If we really want to achieve the goals of this new subject, the assessment design has to be changed accordingly.

3. Why School Based Assessment?

Assessment can include any activity that serves to provide feedback to the teacher and students about what the students has learned (Wilson, 2005). In Liberal Studies, besides traditional examination, teachers can use essay, speech, project, performance task, etc. as SBA components. The more variety used in assessment, the more complete the picture of what the students know and can do (Wilson, 2005). The assessment practices should be aligned with curriculum planning, teaching progression, student abilities and the local school context. Students can receive timely and helpful feedback (Black, 1993) to promote learning as the assessment is continuous and formative in nature. The information collected will help to motivate, promote and monitor student learning, and will also help teachers to find ways of promoting more effective teaching and learning.

Generic skills are fundamental to helping students learn how to learn. The skills transcend specific subjects, disciplines, or applications. One of the aims of Liberal Studies is to develop in students a range of skills for life-long learning, including critical thinking skills, creativity, problem-solving skills, communication skills and information technology skills (EDB, 2007). Skills such as IT skills are easier to assess, but skills such as critical thinking skills and creativity are harder to measure. When we want to do assessment, care must be taken in explicitly defining the learning outcomes associated with generic skills. Some consensus should be developed between teachers who will provide instruction and engage in assessment of the same generic skills in the context of different subjects. Multiple assessment methods in multiple contexts should be used to ensure that the skills are in fact transferable. In the traditional ways of assessment, paper tests and essays are used. However, many generic skills are less susceptible to assessment using these methods. The introduction of SBA could be a good alternative. Teachers can consider using portfolios, projects, role play, demonstrations, presentations, etc. as parts of the SBA coursework. These are good approaches to gather evidence of learning outcomes and are more consistent with assessment for learning. Also, learners can be engaged actively in assessing their own learning. The drawbacks can be labour-intensive for students and teachers, and there might be questions about reliability.

Arter and McTighe (2001) point out that teachers at times are nervous about grading certain types of student work – like creative writing, a research report, or an oral presentation. They are even more anxious when they need to ensure that the students “will think critically” or will be “life-long learners”. The Liberal Studies teachers in Hong Kong are facing similar problems. In order to make reliable judgments on students’ competency in generic skills, standards, criteria and indicators are needed. Some sort of rubric is required to guide the judgment to the evidence of student achievement. Using rubrics in assessment saves time, provides timely and meaningful feedback for students, encourages critical thinking and maintains equity and fairness (Stevens & Levi, 2005). Also, rubrics provide valid and reliable assessment of student learning on the complex and hard-to-assess student outcomes.

In Hong Kong, many teachers do not have the skills and experience in instructing and assessing generic skills. It seems that there is a perceived conflict with traditional curriculum and instructional methods. The use of assessment tools such as rubric and portfolio is not common. Well-established assessment procedures and tools are not adequate at the moment. More teacher development is needed to make teachers better prepared for the new challenge.

According to the Hong Kong Examination and Assessment Authority (HKEAA), School Based Assessment is an assessment carried out by schools as part of the teaching and learning process, with students being assessed by their subject teachers and marks awarded will count towards the results of public examinations. The aims of SBA of Liberal Studies are:

- To enhance self-initiated research and organizing ability
- To enhance various generic skills

One of the disadvantages of public exam is that it only gives a snapshot or a limited sample of student’s performance. Examination is summative in nature and also not very effective in assessing some of the generic skills. In contrast, SBA is more holistic and provides a more comprehensive appraisal of students’ performance. Students’ results are not determined by one single exam. Teachers can give feedback to enhance students’ understanding of their weaknesses and strengths. Students are encouraged to learn independently. This ongoing assessment process can be used to recognize needs, provide assistance with skills, clarify directions, motivate, and guide instruction (Chapman & King, 2005). Black and Wiliam (1999) also suggest that teacher assessment which diagnoses students’ difficulties and provides constructive feedback leads to significant learning gains. SBA can be used to complement public examinations. Teachers will be able to see students’ continuous progress through working on a SBA project or Independent Enquiry Study (IES) over the three year period. According to the Liberal Studies: Curriculum and Assessment Guides (Secondary 4-6) (Final version March 2007) by Education Bureau:

IES is an integral component in the curriculum of Liberal Studies. It is designed to provide a self-directed learning experience which requires students to shoulder the major responsibility of learning and demonstrate self-management skills in pursuing an investigative study on a self-chosen topic, with teachers and other peer students as partners in

learning.

Shepard (2005) points out that the aim of practicing formative assessment is to establish classroom practices that encourage peer assessment, regard errors as opportunities for learning, and promote shared thinking. That means we need to foster a learning culture: classrooms in which both students and teachers focus on learning rather than on grades. However in Hong Kong, when compared to many western countries, still only a small percentage of students can get a place in universities. Harsh competition among students is inevitable. In Australia, UK and USA, 82%, 51% and 64% of the students respectively can receive government subsidized higher education, but the figure for Hong Kong is only 18%. To make things worse, the student population has decreased rapidly in recent years and that means some schools are under the threats of closing down. Schools need good results to attract students. Therefore for some teachers, assessment for learning is fine, but helping students to get good grades is a matter of life and death. As Liberal Studies is entering the stage as a core compulsory subject and SBA is an important part of the assessment, many Hong Kong educators and students are concerning about the fairness of SBA and would like to understand how the system of moderation works.

4. Moderating SBA coursework scores

With the appropriate alignment with assessment criteria and learning objectives, a set of rubrics and guidelines can be produced for teachers to follow to give similar conclusions about student's work (Chase, 1999). Many students' abilities not easy to be assessed by examinations can now be evaluated through SBA. The validity of assessments is therefore improved as the assessment is measuring what it sets out to measure (Weeden, Winter & Broadfoot, 2002). Still, teachers, students and parents, are worry about the fairness of marking. Actually, teachers are best placed to judge their students' performance. They can reliably assess the performance of all students within the school in a given subject. However, when teachers make these judgments, they are not necessarily aware of the standards of performance across all other schools (HKEAA, 2007). Despite training in carrying out SBA, and even given that teachers will assess students on the same task and using the same assessment criteria, teachers in one school may be harsher or more lenient in their judgments than teachers in other schools. They may also tend to use a narrower or wider range of marks. To answer that, a moderation process will be carried out by the HKEAA to ensure the fairness of SBA. In fact, Australia and New Zealand have been using moderation methods for years. According to the New Zealand Qualification Authority:

The main function of moderation activities is to ensure that different applications of standards remain within acceptable limits.

The Board of Studies in Victoria, Australia says:

Moderation is the process of ensuring that the same assessment standards are applied to students from every school doing a particular study.

In Hong Kong, the HKEAA is intending to use moderation methods to moderate SBA scores for Liberal Studies and also other subjects with SBA components, which will be examined for the first time in 2012. According to the HKEAA, the main reason for having moderation is to ensure that SBA results are comparable and fair for all students from different schools.

The commonly used moderation methods include statistical moderation, consensus moderation and expert moderation. Each method has its advantages.

In consensus moderation, groups of teachers conduct moderation meeting occasionally. The teachers select samples of student work from the same activity and brought these to the meetings. The moderation participants then scored all the samples of student work, shared their scores with the group, and then discussed variations in scoring to reach consensus. Teachers also used this to discuss instructional implications as well as reflect on their own practice.

In the state of Victoria in Australia, consensus moderation was used to moderate students' coursework. The study by Ingvarson (1990) (as cited in Roberts and Wilson, 1998) found that involvement in the consensus moderation process:

- added significantly to teachers' skills for assessing student learning;
- enhanced teachers' ability to evaluate and improve their teaching;
- significantly increased teachers' access to useful ideas for teaching;
- enhanced the quality of learning of students;
- affected positively participants' teaching in non-project classes; and
- supported, rather than intimidated, beginning teachers.

Ingvarson (1990) also reported that the positive responses increased as teachers had more experience with moderation, which again reflects the need for teachers to have time to become knowledgeable and skilled in using this process.

The procedures of expert moderation are similar to consensus moderation except that the moderation participants are experts of particular fields. For example, if the topic of the coursework is pollution, experts in environmental protection can give their professional judgment on the quality and depth of the students' work.

The above two moderation methods are non-statistical methods or may refer as social moderation. One of the drawbacks is that consistency may still be at risk even though many moderation meetings were conducted. Different groups might reach different conclusions especially when the unit standards lack specificity or moderation participants lack experience and poses dominant personalities (New Zealand Qualification Authority, 1992). Although there are ways to address the potential problem, a great deal of human effort and time will have to be spent. For example, a representative from each local group meets in panels organized on a national basis and therefore the feedback can be used to improve consistency.

In 2000, the coursework assessment of Victorian Certificate of Education (VCE) in Australia started to include statistical moderation in the moderation process. In Hong Kong, the HKEAA will adopt the statistical moderation method in Liberal Studies School Based Assessment. Statistical moderation is a fair and impartial way of ensuring consistency of standards across schools, while maintaining the students' rank order given by the school (HKEAA, 2007; Board of Studies in Victoria, 1999). This method is particularly appropriate in situations in which there is another measure available that can be used to moderate schools' assessment. Assuming that the public examination result is a valid measure of overall level of performance of students, statistical moderation uses students' exam results to determine if any adjustment in SBA scores is needed. This method is fast, objective and less costly, but the pre-requisite is the assumption mentioned above must be valid.

A school teacher may be harsher or more lenient than teachers in other schools. The spread of students' scores given by teachers may be different as well. Statistical moderation tries to solve the two problems by adjusting the average and the spread of SBA scores of students in a given school so that they are aligned with the exam scores of the same group of students. So the mean of SBA scores may be adjusted upwards or downwards and the deviation of scores may be stretched or compressed. The SBA scores will then be comparable across various schools while the rank ordering of SBA scores within a school remains unchanged. The statistical moderation method to be employed by the HKEAA will involve eight steps. The following is summarized from "Statistical Moderation of School-based Assessment Scores, HKEAA (2007)".

Step 1: Identify moderation group.

The group is defined by all students studying Liberal Studies taught by the same teacher.

Step 2: Convert the raw scores to a scale of 100 points.

The conversion is for the purpose of calculation and comparison.

Step 3: Calculate moderation variable.

Liberal Studies examination consists of two papers. The weighted combination of the exam scores will be the moderating variable.

Step 4: Derive moderation formula by statistical methods.

Here, outliers are excluded to avoid distortion. A candidate's moderated SBA score is a function of the predicted SBA score of the group, multiplied by a factor to take into account the relative spread of the SBA scores on the exam. The formula is like this:

$$F(x) = x_{mean} + \beta(\bar{z} - z_{mean}) + (x - \bar{x}) \frac{s_p}{s_x}$$

where x = the SBA score of candidate

x_{mean} = the global mean of the SBA scores

z_{mean} = the global mean of the moderator variable (exam) scores

$\bar{x}$ = the group mean of the SBA scores

$\bar{z}$ = the group mean of the moderator variable (exam) scores

s_x = the group standard deviation of the SBA scores

s_z = the group standard deviation of the moderator variable (exam) scores

$s_p = \sqrt{\frac{(s_x^2 + s_z^2)}{2}}$ or the mixed group standard deviation

β = the slope calculated from regressing SBA scores on moderator variable scores, taking

into account the nesting of candidates within group

Step 5: Calculate moderated SBA scores.

Step 6: Convert moderated SBA scores back to the original SBA metric.

Step 7: Post hoc checking.

To identify outliers for individual review.

Step 8: Combine moderated SBA score with exam scores to form total subject score.

After statistical moderation, the HKEAA will send an SBA moderation report to each individual school for reference. The details include the statistics of the moderation group and comments:

Statistics of the moderation group:

Moderation group ID

Mean of SBA scores

Standard deviation of SBA scores

Mean of moderated SBA scores

Standard deviation of moderated SBA scores

Number of candidates in the moderation group

Comments:

The mean of the SBA scores is (within the expected range / much higher / higher / slightly higher / slightly lower / lower / much lower than expected)

The spread of the SBA scores is (within the expected range / wider / slightly wider / slightly narrower / narrower than expected)

The report can provide useful feedback to teachers and schools to understand if they were being too harsh or the spread of score is too narrow and vice versa.

School scores may be adjusted up or down after statistical moderation and the rank order of the students remains unchanged. Individual candidates' SBA scores are not directly affected by how well they performed as individuals but as a group in the exams. To achieve the highest possible final score, a student will need to achieve the highest possible SBA score and the highest possible exam scores. This method seems fair and scientific. It can be carried out in a very short time using very few resources. At the same time, teachers' judgments are valued in terms of preserving the rank order of students.

However, there are issues needed to be addressed. Firstly, statistical moderation relies on the assumption that the public exam is a valid measure of the overall performance of the moderating group. This assumption is valid in most academic subjects. But, what about Liberal Studies? It could be possible that the exams and SBA are measuring different things. If the whole assumption is invalid, the whole statistical moderation has little use. The objective of Liberal Studies coursework is to enhance self-initiated research, organizing ability and various generic skills (HKEAA, 2007). So SBA is to measure the above skills. The public exam claims to measure the overall performance of students. Therefore, it must at least include some items to assess students' research, organizing and generic skills. Even though we can design such items for written exams, they may not be most effective and reliable. Perhaps the abilities measured in the exams are highly correlated to the abilities measured in SBA. Therefore, the exams need not cover what SBA is supposed to assess, and can provide an accurate prediction. But we need evidence for that. Also, if that is the case, a paradox appears. If we can use public exam to measure what SBA is measuring, then why not just use public exam? And yet, we are using SBA to measure the skills public exam cannot measure, as said by the HKEAA (2007): "Certain components of some curricula cannot be assessed within the context of a written examination, and this can be complemented by SBA." As we can see, the exam and the SBA are complementing each other. They are probably not measuring the same things. The New Zealand Qualifications Authority (1992) comments on the major technical difficulties with the method: "The dissimilarities between the elements assessed in the test and those outside the test may render the scaled results invalid."

In statistical moderation, there is no human intervention. This could be a good thing as it eliminates subjectivity. However, before the moderation process begins, the rank order of a moderating group is solely determined by one single teacher and nobody else. Although teachers are professionals with professional ethics, they might still have their preferences. Students may present their work in very distinct and sometimes creative formats. Their work can demonstrate the range and depth of a student's learning, knowledge and skills over time and across different contexts.

Students are actively involved in the selection of materials which provides opportunities for reflection and gives them a sense of ownership. It is unavoidable that different teachers may favour different ideas. Reliability refers to the consistency of assessment results (Gronlund, 1998). It is possible that the rank order of a group may be different if it is determined by other teachers. Hence, the reliability of teachers' marking may be at risk as there are no moderation meetings, no external assessors and no common assessment tasks. The problem may escalate if the number of students in the group is small.

Another drawback is that there is little chance for professional development and sharing of ideas which can provide invaluable feedback to teaching and learning. Just providing the moderation report to schools is probably not as effective as teachers of different schools discussing and marking coursework together.

The HKEAA should use Australia's experiences as reference. On top of statistical moderation, the Board of Studies in Victoria, Australia uses cross-marking and common coursework tasks to increase the comparability of school assessment. Cross-marking of carefully selected samples of student work can help teachers to establish a common application of the assessment criteria and marking scheme for each task. Each teacher can then ascertain the extent to which the marking standards need to be adjusted to align with the common standard. Since the coursework tasks set by the teachers may vary and the ways students present their work may be very different, choosing a few tasks from a pool of common coursework tasks can help to achieve comparability.

Any adjustment to students' score is determined by the exam scores for the whole group, not by the student's own exam score. Therefore, some people concern that high-performing students will have their scores reduced simply because they are in a low-performing group. In Australia, the moderation procedures used can eliminate that at the top end of the scale by setting the highest moderated score to be equal to the highest exam score. The procedures can also eliminate the reverse situation at the bottom end of the scale. However, the HKEAA does not have such procedures. Rather some outliers are temporarily excluded from calculation to avoid distorting the moderation results. In figure skating competition, the bottom two and top two scores are eliminated and the remaining eight scores are then averaged. This is based on the assumption that the high and low scores are biased. But a biased judge could have given biased marks throughout. The critical question is the judge's consistency in interpreting the scoring system consistently all the time. The Liberal Studies SBA will also face the problem of identifying biased markers. The use of Item Response Theory and Many-Faceted Rasch Model could be the answer. The model uses shared marking to establishing consistent severity levels of markers. Then Rasch Model is used to determine the markers severity level. As linked data (items are marked by more than one markers to form links) are necessary, the judging plan needs to be carefully designed. The Many-Faceted Rasch Model is simplified as follows:

$$\text{Log}(\text{Pr}(\text{correct response})/\text{Pr}(\text{incorrect response})) = B_i - C_j - D_k$$

or

$$\text{Pr}(\text{correct response}) = \frac{e^{B_i - C_j - D_k}}{1 + e^{B_i - C_j - D_k}}$$

where B_i is the ability of the i^{th} student

C_j is the difficulty of the j^{th} item

D_k is the severity of the k^{th} marker

In here, there are three facets or three factors (B_i, C_j, D_k) which may affect the probability of getting a correct response. After data input and calculation using the software "Facets", the model will then give us outputs such as marker consistency, bias analysis and marker discrimination. Markers who were too lenient or too harsh are identified when the data and the model does not fit. Follow up actions such as re-training, cross-marking and double marking can then be taken.

Over the past 20 years, many examination bodies have adopted SBA to improve the quality of teaching, learning and assessment. In Hong Kong, the weighting given to Liberal Studies SBA is only 20%. In Australia and UK, the weighting is about 50%. Provinces in Canada have up to 100% SBA and the International Baccalaureate (IB) has a heavy emphasis on SBA supported by teacher moderation arrangements. Chan and Cheung (2007) point out that Liberal Studies aims at helping students to acquire important skills, knowledge and work habits that cannot readily be assessed or promoted through paper-and-pencil testing. The outcomes are essential to learning within the disciplines and also valued by tertiary institutions and employers. Furthermore, the SBA activities are probably more enjoyable and meaningful to students. Therefore if we consider the nature of Liberal Studies and overseas experiences, giving SBA a 20% weighting may be too conservative. But why people have reservation? Survey results indicate that schools and teachers have divergent views as to whether SBA is doable and whether the effort required by teachers and students is reasonable (Chan & Cheung, 2007). From the responses of many schools that they believe they are expected to do all things they have done in the past as well as carry out SBA as an extra responsibility. Uncertainties and increase in

workload are teachers' main concerns. The solution to the overloading problem is to integrate SBA with teaching programmes so that SBA will become a natural part of daily work and replace some of the traditional assessment practices within schools. All schools will start offering Liberal Studies in 2009. But not all teachers have the experience in implementing SBA. The HKEAA has to organize professional development programmes and provide exemplars and detailed guidelines to strengthen teachers' understanding of SBA and to ensure the consistency of teachers' assessment. Fortunately, the HKEAA has promised to publish detailed guidelines on how to implement Liberal Studies SBA, and also provide training for future Liberal Studies teachers.

5. Conclusion

The education reform Hong Kong currently undergoing composes of many initiatives. The introduction of Liberal Studies is only a small part of it. Nevertheless, it is affecting the education community. The traditional mode of teaching, learning and assessment is no longer adequate. Teaching without learning is a worthless exercise, and hence the assessment of whether students have learned is integral to the process of education. Teaching is a process to obtain our goal of learning, as the desired outcome. Assessment is the process used to determine whether the outcome was met (Wilson, 2005). The learning goals of Liberal Studies include developing many skills difficult to be assessed by written exams. Once the problems of fairness and overloading are solved, SBA could be a suitable platform for assessing generic and thinking skills as well as practicing learning-oriented assessment.

References

- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Corwin Press, Inc.
- Bedford, J. (2002). *Washback – the Effect of Assessment on ESOL Teaching and Learning*. http://www.tki.org.nz/t/esol/esolonline/secondary_mainstream/research/nzresearch/jenni_bedford/jenni_bedford.pdf [Accessed on 8th July 2007].
- Black, P. (1993). Formative and summative assessment by teachers. In *Studies in Science Education*, 21, 49-97.
- Black, P. & Wiliam, D. (1999). *Assessment for learning: Beyond the Black Box*. Cambridge.
- Board of Studies, Victoria, Australia (1999). *Statistical moderation of VCE coursework*.
- Chapman, C. & King, R. (2005). *Differentiated assessment strategies. One tool doesn't fit all*. Corwin Press.
- Chan, K. K., & Cheung, F. (2007). *School-based assessment: Changing the assessment culture*. HKEAA.
- Chase, C. I. (1999). *Contemporary assessment for educators*. Longman.
- Cheng, Y. C., Chow, K. W. & Mok, M. M. C. (2004) (Eds.). *Reform of teacher education in Asia-Pacific in the New Millennium. Trends and challenges*. Kluwer Academic Publishers.
- Cheung, W. K. & Hui, P. K. (2006) *Liberal Studies*. <http://www.ln.edu.hk/mcsln/liberal14.html> [Accessed on 20th July 2007].
- Education Bureau (jointly prepared by the Curriculum Development Council and the Hong Kong Examinations and Assessment Authority) (2007). *Liberal Studies: Curriculum and Assessment Guides (Secondary 4-6) (Final version) (March 2007)*.
- Education Bureau (2004). *Reforming the Academic Structure for Senior Secondary Education and Higher Education - Actions for Investing in the Future*.
- English, F. W. (2000) *Deciding what to teach and test. Developing, aligning and auditing the curriculum*. Corwin Press, Inc.
- Glaserfeld, E. V. (1989). Learning as a constructive activity. In *Developments in learning and assessment*. P. Murphy and B. Moon (eds.). The Open University.
- Gronlund, N. E. (1998). *Assessment of student achievement. 6th edition*. Allyn & Bacon.
- Hong Kong Examination and Assessment Authority (HKEAA) (2007). *Statistical Moderation of School-based Assessment Scores*.
- New Zealand Qualifications Authority (1992). *Designing a moderation system*.
- Roberts, L. & Wilson, M. (1998). *An Integrated Assessment System as a Medium for Teacher Change and the Organizational Factors that Mediate Science Teachers' Professional Development*. University of California, Berkeley.
- Shepard, L. A. (1997). *Measuring achievement: What does it mean to test for robust understanding?* William H. Angoff Memorial Lecture Series. Princeton, NJ: Educational Testing Service.
- Shepard, L. A. (2005). Link Formative Assessment to Scaffolding. *Educational Leadership / November 2005*.
- Stevens, D. D. & Levi, A. (2005) *Introduction to Rubrics*. Stylus Publishing.

Weeden, P., Winter, J. & Broadfoot, P. (2002) *Assessment. What's in it for schools?* RoutledgeFalmer.

Wentworth, J., & Davis, J. R. (2002) Enhancing interdisciplinarity through team teaching. In C. Haynes (ed) (2002) *Innovations in interdisciplinary teaching*. The Oryx Press.

Wilson, L. W. (2005) *What every teacher needs to know about assessment*. 2nd edition. Eye On Education Inc.

Table 1.

Component	Part	Weighting	Duration
Public examination	Paper 1	50%	2 hours
	Data-response questions	30%	1 hour and 15 minutes
	Paper 2 Extended-response questions		
School-based Assessment (SBA)	Independent Enquiry Study (IES)	20%	