

This is the final version of this Chapter prior to acceptance and publication: July 2015

Implementing a learning-oriented approach within English Language assessment in Hong Kong schools: practices, issues and complexities

Liz Hamp-Lyons

University of Bedfordshire, UK

ABSTRACT

From 2007, the education system in Hong Kong introduced a school-based assessment component to its English Language examinations at age 15-16/16+. Based in the principles of assessment for learning (AfL) (Black & Wiliam 1998) this was a major educational shift for the system, the schools, and especially for the teachers. This innovation at first met with great resistance, but is now 'bedded in'. Nevertheless, it continues to be less than fully understood by its stakeholders and by the education society in Hong Kong as a whole, including teacher-educators. The paper places attempts at assessment innovation in the context of Hong Kong, reprises the aims and structure of the school-based assessment (SBA), and describes some of the challenges of development and implementation. It goes on to consider some of the cultural and political influences that helped and hindered the effective implementation of this remarkably different form of assessment in Hong Kong, and argues for increased attention to (language) assessment literacy in Hong Kong, and wherever assessment innovation is to be attempted.

Key words: language assessment literacy; learning-oriented assessment; school-based assessment (SBA); Hong Kong

Author biodata:

Liz Hamp-Lyons held professorial positions at universities in Hong Kong and then at the University of Melbourne, before joining the Centre for Research in English Language Learning and Assessment, University of Bedfordshire, UK. She is Guest Professor at Shanghai Jiao Tong University, where she is the senior consultant to the College English Test (CET) and the Profile of Advanced Competence in English (PACE). Her research interests include the development and validation of English language performance (i.e. writing and speaking) assessments, assessment for academic and specific purposes, assessment for learning, learning-oriented language assessment and language teacher assessment literacy. Recent funded projects have researched school-based speaking assessment, writing assessment for professional purposes, and the learning opportunities of the Cambridge English: First speaking test. She was founding Editor of the *Journal of English for Academic Purposes* and current Joint Editor in Chief, and Editor of *Assessing Writing*.

1. BACKGROUND: EDUCATIONAL AND ASSESSMENT INNOVATION IN HONG KONG

It is becoming increasingly well understood that every educational innovation thrives or flounders within a social-political ideological context (Henrichsen 1989; Kellaghan & Greaney 1992; Wall 2005). Hong Kong has for many years had a traditional norm-referenced examination system for school placement, promotion and exit (for a historical overview of the public examination system in Hong Kong, see Choi & Lee, 2008). While this system is congruent with a traditional Chinese cultural heritage context, educators have long felt there is something fundamentally flawed about a system in which students may fail every school subject in which they take a formal exam. In the English subject, for example (by no means one with the worst results), between 1997 and 2007 (the last year of norm-referenced results reporting) 41-60% of students failed the Syllabus A English and 60-78% failed the more difficult Syllabus B English. Steps have been taken at several points in the past 30 years to reform the educational system to better fit the needs of school students, and to ensure that the right individuals enter tertiary education and that sound educational opportunities are available to those not entering tertiary education (King 1994; Qian 2008). The need for a more liberal and broad approach to examinations was one of the points made in the 1997 Education Commission Report, *Quality School Education*: this Report may well have had foreknowledge of one of the recommendations of a review of the public examination system which was at that time reaching a conclusion. The review, which has become referred to as the *ROPES Report* (*Review of public examination system*) was not made public, but the general outlines of its recommendations became known. Among the members of the review consultancy team was Patricia Broadfoot from the UK, well-known for her progressive views on inclusive and humanistic assessment and so it is not surprising that one comment made in the review was that worldwide there is “a pronounced shift in responsibility for assessment of student achievement to a blending of the information available from both (traditional and classroom) sources” (as reported by Berry 2008). However, as Fok, Kennedy, Chan & Yu (2006) comment: “Hong Kong is famous for its examination-dominated culture, which heavily relies on public examinations. So ingrained has it become that the whole society is sensitive to any change in such an assessment mechanism” (p.1).

One recommendation of the ROPES Report was to expand school-based assessment in the Hong Kong Certificate of Education (HKCE) and Hong Kong Advanced Level (HKAL) Examinations. Christina Lee, who joined the Hong Kong Examination Authority (HKEA) in 1990 as English Language Subject Officer and was the General Manager of the Assessment Development Division, describes this history:

“HKEAA has had something what we called ‘teachers assessment scheme’ (TAS) as early as 1978, so we didn’t call it ... SBA..., but in effect, they are like the same thing, I would say, you know, the teachers assessment scheme and SBA, but, basically, when I started working with the exam authority in the early 1990s, TAS was already going on in

Chemistry, Biology, and then later on Physics also joined.” (C. Lee personal interview, May 16, 2007)

Yung (2002) describes the form of ‘school-based assessment’ implemented in the early years of reform as comprised mainly of fairly traditional formative assessment, in which essay marking is thought of as feedback as well as ‘traditional’ assessment; and in which ‘school-based’ referred primarily to a de-centralisation of some components of assessment to the schools. In 1999 the Hong Kong Education and Manpower Bureau, the governmental body with oversight of the HKEA, commissioned its own ‘Strategic Review of the HEAA’ which was published in 2003 (IBM Corporation, 2003). While continuing the direction of the 1998 review of the public examination system, this was a very ambitious and progressive review with far-ranging recommendations, aiming simultaneously at financial and managerial reform of the Authority, and at reforming the testing culture in Hong Kong to be less rigid and more supportive of modern thinking in curriculum, teaching and learning. The re-naming of the HKEA as the HKEAA (Hong Kong Examinations AND ASSESSMENT Authority) was a direct result of this review. One of the most immediate recommendations of the Strategic Review was to “progressively move to the further development of school-based assessment... and to assessment based on defined standards” (p.100). But the Review also acknowledged that “... Hong Kong must urgently engage in fundamental debate right across its education community – and with the public – to raise ‘assessment literacy’”. (p. 3). In fact, the consultants’ consciousness of the complexity of educational and assessment reform and innovation permeates their review. This recommendation was accepted by the Hong Kong Legislative Council in December 2003 [LC Paper No.CB(2)634/03-04(01)] and funding was earmarked. As the innovation was gradually introduced in various school subjects, responses from teachers and parents were mixed (Cheung 2001).

2. AIMS AND STRUCTURE OF THE HONG KONG SBA IN ENGLISH LANGUAGE

2.1. Aims and first steps

In 2004 the HKEAA put out a call for tenders for the development of a new component to the Hong Kong Certificate of Education Examination (HKCEE) English Language syllabus which would be a school-based assessment (SBA) component of speaking, and would first be introduced to secondary forms 4 and 5 in the 2005-6 school year, and included in exam results reported at the end of the 2007 school year. The purpose of this innovation from the government viewpoint was to align assessment more closely with developments in the English Language teaching syllabus and the Senior Secondary curriculum (Curriculum Development Council, 1999). The new syllabus, set by the central Curriculum Development Institute of the Hong Kong Education Department, was to include a speaking component taught and assessed in the classroom by the teacher, which would contribute 15 percent of the student’s total English grade. A proposal in response to an invitation to tender for the project was submitted to the HKEAA by a team

in the Faculty of Education at Hong Kong University (HKU), with Dr. Chris Davison as Project Leader and Prof. Liz Hamp-Lyons as Principal Researcher in September 2004. The proposal was for a pilot study with 10-12 schools in the first year; however, the Hong Kong Education Bureau (EDB) were concerned that not to include all schools would be seen as unfair, and determined that this innovation should be introduced full-scale. The proposal was therefore revised in December 2004 and awarded on December 24th 2004.

2.2. Early challenges

Work on the design and development of this new approach, to be called School-based Assessment (SBA) – English, began in January 2005 with a deadline of August 2005 for full functionality. The development team was asked to present an overview of and rationale for this new approach in March 2005, to a live audience of about 1,000 teachers. The plan from the Educational and Manpower Bureau (EMB) was for it to be rolled out in September of that year with all secondary schools in Hong Kong (a little over 500) without prior trialling or a familiarization period. This sudden and rapid change in assessment coupled with the inevitable (and essential) changes in teachers' teaching strategies caused anxiety among school principals, parents, and especially teachers, and attracted considerable media attention. This negative attention led to intense scrutiny and discussion at senior education levels, and in January 2007 a statement was issued by Dr K K Chan, the Principal Assistant Secretary (Curriculum Development) in the Hong Kong Government's EMB and Dr Francis Cheung, then Deputy Secretary General of the HKEAA, strongly supporting the introduction of the SBA and offering consultation on implementation. As a consequence, the plan was changed and every school was given the choice of introducing the new English SBA on the original timetable (i.e., formal reporting at the end of the 2006-7 school year), or of waiting one year before introduction, or waiting two years before introduction. This staged introduction of the assessment innovation enabled some interesting and informative research, which the HKU consultancy team also conducted for the HKEAA, and which was reported as two 'longitudinal studies', the earlier of which can be found on the HKEAA website: [http://www.hkeaa.edu.hk/DocLibrary/Resources/Longitudinal_Study-SBA_HKCEE_English_dec2010.pdf].

2.3. Structure of the Hong Kong SBA in English Language

The Hong Kong School-based Assessment is tied to the same standards-referenced assessment approach as the rest of the public exam system. However, as the Deputy Secretary General of the HKEAA, Dr. Pook, says on the HKEAA website: [http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/sba_definition.html], the SBA component of the English Language exam aims to go further by giving a more comprehensive appraisal of learners' skills and abilities, including aspects that cannot be easily assessed in traditional large scale examinations. The key audience for the DVDs and web-based materials for the English language SBA are teachers, and at [http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/atl_interrelationship.html] Liz Hamp-Lyons explains the interrelationships between assessment, teaching and

learning, and Dylan Wiliam discusses the differences between formative and summative assessment.

The major components and requirements of SBA in HKCE English Language are described in Davison and Hamp-Lyons (2010), Lee (2008) and Davison (2007), and are summarized below:

- Students are assessed on 4 domains of speaking custom-designed to fit the two speaking task-types: individual presentation and group interaction.
- Students are assessed during class time by their own English teacher.
- The assessment is embedded into the curriculum.
- Teachers video or audio record a range of the student assessments they observe and assess, in order to assist with later standardization.
- Teachers submit the best two sets of scores out of three assessment tasks.
- SBA speaking scores constitute 15% of each student's total HKCE English Language result.
- The content for students' speaking is drawn from an extensive reading/viewing programme using four different text types: print fiction, non-print fiction, print non-fiction and non-print non-fiction.

This reading/viewing-into-speaking process is closely similar to a curriculum sequence familiar in both English and other disciplines, where students read or work with other input materials, meet in groups to discuss it and expand their individual understanding by the sharing of viewpoints with others, and the *group interaction task* aims to model such sharing. The *individual speaking task* is seemingly less 'authentic', but within classroom instruction the individual report-back is quite a common task, as is the more formal individual presentation in the senior secondary years as well as in the universities, in many disciplines. These task types and parameters were established by the consultancy team during extensive consultations with a core group of volunteer teachers from a group of over 50 schools who participated in seminars and workshops, action research in their own classrooms, and the trialling and validation processes of the SBA English Language as it developed. Some of these requirements were predetermined by the HKEAA in order to comply with the curriculum set by the Hong Kong Curriculum Development Council (CDC).

These tasks, the individual presentation (IP) and the group interaction (GI), will not sound very unusual or innovative to most readers, and indeed they are not, and they probably suggest speaking activities that are quite inauthentic, as Lam (this volume) suggests. However, it should be remembered that in Hong Kong there had in 2005 been little overt *teaching* of speaking skills, and that the *assessment* of speaking is itself quite new in Hong Kong (Andrews & Fullilove 1994). The formal speaking test that was introduced in 1994 was a close imitation of speaking tests used at that time in the UK, and was (and remains) extremely formal and narrow, and was not very reliable. Nevertheless, Andrews and Fullilove (1994) argue that the sheer fact that speaking was now officially assessed played a part in bringing teachers' and schools' attention to the teaching of speaking.

However, what was innovative in Hong Kong in 2005-6 was the expectation that classroom teachers would design and carry out these task types in their own classrooms, and that they would be allowed to score their own students' performances. We quickly found that most teachers needed strong support in developing appropriate reading-into-speaking tasks for their students so that they would have content to work with in fulfilling a task. An IP task might be, for example, describing an interesting character in a book they have read or a film they have viewed. Teachers asked for guidance on how to plan the tasks and how to tie the assessable task to the curriculum. Davison (2007) shows an exemplar IP task taken from the early stages of the materials development. The GI task is more complex, as described by Gan, Davison and Hamp-Lyons (2009), being a dialogue or exchange of short turns between two or more speakers. Turns are expected to be comparatively short and quite informal. A student taking part in a GI needs to attend to turn-taking skills, and to be able to initiate, maintain and control the interaction through suggestions, questions and expansion of ideas. These interactive skills are deliberately mentioned and rewarded in the assessment domains, criteria and levels. The expectations of effective student performance on the GI task also include the capacity to speak intelligibly and reasonably fluently with suitable intonation, volume and stress, using pauses and body language such as eye contact appropriately and effectively; to use a range of vocabulary and language patterns that are accurate and varied, and language that is natural and interactive, not memorised or read aloud. The Guidelines also note that some use of formulaic language may appear when appropriate for structuring, but that overuse of set phrases is discouraged. The assessment criteria for the GI can be found at http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Assessment_Criteria.pdf

2.4. Expansion into the new 3-3-4 senior education structure: SBA 2009-present

By the time the SBA for English Language in the HKCEE reported its first results in 2007, the plans to change Hong Kong's senior secondary and university structure from 3-4-4 to 3-3-4 (i.e., from 4 years of senior secondary school and 3 years of university to 3 years of senior secondary and 4 years of university—or, from the British to the American system) were already well advanced. As a consequence of this major restructuring, the HKCEE was replaced and the Hong Kong Diploma of Secondary Education (HKDSE) was introduced from September 2009. The HKEAA mandated that in the HKDSE the SBA English Language should be revised and extended to cover the full three years of senior secondary education, and in the third year there should be the opportunity for students to use SBA within their chosen elective subjects. The original consultancy team were asked to revise and expand the existing materials for teachers to fit the new requirements. The HKEAA provided support for further materials development and for extension and renewal of the professional development materials and courses. The revised, 3-year, SBA English materials and additional professional development (PD) materials were introduced in 2009, and PD courses continued to be run until the funding ended in December 2011.

2.5. Seeking balance between summative and formative paradigms

Successfully transitioning schools, teachers, learners and parents into this 'school-based' approach to assessment is much more than a matter of developing and distributing sets of task types and parameters, sample materials, criteria for assessing spoken language in these contexts, and samples of performance. We did all those things. The Guidelines for SBA, which explained the nature of the oral text-types to be assessed, the mandatory assessment conditions, bureaucratic issues such as record-keeping and particularly the importance of standardization, were extensively shared among schools, in seminars run at the Hong Kong University, in schools, and for the teachers' support network HKEdCity [http://www.hkedcity.net/article/project_sba_eng/050902-002/]. The Guidelines were (and are) available online. The team worked with a clear focus on the need to work with teachers to create successful innovation reform. A key aspect in this was 'accountability': not only the education system but perhaps even more, the teachers and school principals wanted to be assured of the trustworthiness and fairness of this new kind of assessment.

Given the tensions of the Hong Kong traditional exam context (Hamp-Lyons 2007), we had to find a balance between the formative thrust of AfL and the summative purposes teachers were accustomed to. This balance, though explicitly argued for in the stated goals of the ROPES review in 1998, and again in the IBM report 2003, and despite the movement towards a modest version of school-based assessment in other subject area (Cheung 2001; Yung 2002) had by no means been achieved before our project began, and as Qian (2014) points out, it has still been only partially achieved. As my colleague Chris Davison has recently commented, "... teachers do not work in isolation; their attitudes and beliefs about change are inextricably linked to those of other members of the educational community, not just colleagues, but supervisors, parents and students" (Davison 2013: 271). In addition to the key concepts and principles of AfL, the professional development programme focussed on helping teachers to unpack and understand the structure of the assessment instrument so that they can not only use it for assessing students in their own classes at the appropriate times, but use or adapt it within the teaching and learning they do during instructional units (Davison & Hamp-Lyons, 2010).

The movement towards AfL begins from the challenge to summative assessment as being inflexible, behaviorist and teacher-directed, and proposes modes of assessment that are more reactive to the needs of contexts and learners, drawing on constructivist views of learning, and emphasizing the role of assessments in helping students learn how to learn. An early draft of the Guidelines, which laid out these principles as well as the processes involved in putting them into practice, was introduced and explained in face to face seminars to over 600 local teachers of English during April and May 2005 (immediately after the plans for English SBA had been made public, but before the system itself had been fully developed). The processes schools and teachers use have remained the same since the implementation of the English Language SBA:

- All teachers teaching with SBA meet within their school, view a range of student performances, consider tasks and criteria, and discuss scores.
- Teachers are not required to alter their own scores after discussion, though they

may do so.

- Groups of schools’ “SBA coordinators” meet once a term to view a range of their own schools’ performances, discuss issues within SBA [not only scores] and refresh their shared understanding of what the levels look like (anchor sets are available).
- Scores are not changed, the discussion is advisory.
- Scores from each school are reported to HKEAA, which conducts a statistical “moderation” that may or may not lead to score adjustments.

This very simple description of a process overlays a complex set of procedures that were carefully developed with the aim of achieving two opposing goals: to satisfy the HKEAA and the Hong Kong Education Bureau’s desire for rigorous safeguards for reliability and fairness; and to encourage in teachers the kinds of flexible handling of student learning in live classrooms implicit in the *assessment for learning* movement. These Guidelines were revised twice more before the first teachers carried out the first implementation of a formal (i.e., scored and recorded) SBA teaching and assessment sequence with their classes, and they have been revised several times between 2007 and 2012. It was, and is, regrettable that in their formulation in a simple format intended to make the SBA process seem non-intimidating to teachers, they acquired a bureaucratic flavour and a less than humanistic tone that sits uneasily beside the very humanistic aims of the innovation.

During the initial development stage, as we worked closely with our core group of teachers, they provided feedback on the emerging Guidelines; and as we visited schools and classes and saw these teachers working with SBA, we became more and more convinced of the need for professional development support to teachers working with this very different kind of assessment for the first time. Our collaboration with some of these teachers continued all the way through to the end of the project in 2011. The PD support we developed was of two broad kinds: general language assessment literacy, and specific familiarization/ training and standardization materials.

3. SUPPORTING TEACHERS

3.1. Language assessment literacy in Hong Kong

As Kennedy (2013) has pointed out, increased teacher and student workload, lack of community confidence in school-based processes and even lack of confidence by teachers themselves emerge as key issues during the implementation of SBA in Asian cultures; but he also considers that teachers are reluctant to accept responsibility for high stakes school-based assessment. Carless (2011) has discussed at length the characteristics of examination-driven education in what he calls the ‘Confucian-heritage cultures’ (CHC) of China, Japan, South Korea, Taiwan, Singapore and Hong Kong. All these cultures have been directly or indirectly influenced by the Chinese Imperial civil service examination system, success in which was essential to a respectable post in the civil service and thus, a comfortable life style. Elman (2000) emphasises the competitive nature of the Imperial system and also the draconian regulations surrounding exam-

taking. The Imperial exams focused largely on understanding and interpreting Confucian thought as represented by the key Confucian texts, memorization and exegesis of these texts at a high literary level was most prized. Scholars have argued that this long tradition has led both to the examination-oriented approach to education inherent in these cultures, and also to a view of the reproduction of knowledge as a significant purpose of education. Exams which privilege knowledge-telling rather than knowledge-making lend themselves more easily to cheating.

Hong Kong English teachers have somewhat limited teacher education before they begin teaching, and a substantial proportion of those entering schools as teachers are not qualified, but begin work on the basis that they will qualify during their initial years as teachers. Until very recently, it was quite common for teachers of other subjects to be assigned to teach English because their own English was judged (by the school principal) to be good, or “good enough”. With these varied backgrounds in pedagogy and subject knowledge, teachers bring not just lack of knowledge about assessment but also their own experiences of teaching, learning and assessment in the rigid environment of the typical Hong Kong secondary school. They bring an ingrained belief in external exams as ‘fair’ mechanisms of selection, as well as the belief that it is possible to succeed on any exam with enough hard work. Their students bring these same values, and they are likely to have a little ‘help’ in their own studies and exams from cram schools.

The lack of assessment literacy by teachers is a major stumbling block for any assessment innovation, just as lack of the appropriate knowledge and understanding is a stumbling block for any educational reform. Similar issues have been identified and discussed in China mainland (Chen & Klenowski 2009). Despite a more established structure of English teacher education in Hong Kong, and in spite of the long history of teaching English in elite secondary schools, the situation has been very similar (Boyle & Falvey 1994). This seemingly anomalous situation is in part explained by the social/political history of education and the medium of education in Hong Kong, and is well reported by Evans (2000). ‘Shadow education’, or private tutoring and ‘cram schools’, is very prevalent in Hong Kong (Bray 2007; 2012). Chan and Bray (2014), focusing on Hong Kong in particular, comment that “Much of the shadow education focuses on techniques for performance in external examinations, and is not consistent with the emphases stressed by teachers and the government.” (p.361). It is often argued that reforms in mainstream teaching are undermined by the use of much more traditional methods such as complete dependence on a single textbook custom-written to any new exam syllabus, and the use of read-repeat-recall (rote memorization) techniques in cram school classes. Although there is very little empirical research into the methods and culture of cram schools, one thing we do know is that cram or tutorial schools in Hong Kong may not be officially registered; tutors in cram schools are often not qualified as teachers, and may have BAs in the subject but rarely more advanced subject qualifications; and the charges in these schools as well as the service they provide are uncontrolled. Tutors rarely attend professional development seminars because they are not on the radar, and might not attend anyway as their legal status is questionable. This “shadow education” system therefore influences students’ and parents’ attitudes to every potential educational innovation.

The English Language School-based Assessment was introduced to Hong Kong within the complex context described above, and this reality limited and to some extent shaped what we could do. We were not surprised when teachers expressed concern about this attempt to bring assessment for learning into Hong Kong's English language classrooms, but the opposition of some Hong Kong teacher educators, despite the initiative having been debated and promoted by the government for more than five years, and despite it being officially supported by the universities, did surprise us. Several major articles and letters were published in the Hong Kong English language and Chinese language press, and these were mainly negative/cautionary. For example, the article by Icy Lee (2005), then an assistant professor in education at Hong Kong Baptist University was titled "Latest initiative adds to burden". Lee wrote:

"While most educators applaud the move away from norm-referenced to standards-referenced assessment in the public exam component, the implementation of SBA is causing teachers to shudder. The simple reason is that SBA will put the onus on English teachers to take charge of the whole business of assessing students, and to ensure that the assessment is fair and reliable. The task is daunting, and teachers are ill-equipped for it."

This comment sums up well the sentiments of most teachers at that point. Chris Davison and I met group after group of English teachers, in all of which there were some teachers who spoke up forcefully against the implementation of SBA on the grounds of workload and concern about its fairness. As the project proceeded we learned a great deal about what we could not assume about teachers' assessment knowledge (and also about their pedagogical knowledge), and about what we could not do with the SBA as a classroom-based assessment. We quickly realized that the key to successful introduction of this planned innovation would be the planned provision of the essential aspects of language assessment literacy to teachers who would be teaching with SBA—some 7,000 teachers over more than 500 schools during the first five years, as well as provision of extensive familiarization/training and standardization materials.

3.2. SBA familiarization/training and standardization materials

During 2005-7, substantial work and time went into the development of several kinds of training and familiarization materials. Our core group of teachers helped us by allowing project staff into their schools and classrooms to video record both group interactions (GIs) and individual presentations (IPs): we collected more than 500 such videos. These formed the basis of both the familiarization materials, with professional development activities built around them, and the materials to be used by Group Coordinators (key teachers, mainly taken from the core group, who provided locally-based training and support in 39 sub-districts of Hong Kong). The team also developed a face to face professional development programme which supported and augmented the training materials, and which was delivered a number of times each year between 2009 and 2012, serving a total of almost 8,000 teachers. During these formal courses teachers saw a wide range of video samples and became familiar with the criteria and scales. Every term,

teachers of classes using SBA meet with the SBA coordinator in their own school to discuss video samples from their own students, and to refresh their understanding of the criteria and standards at each scale level by relating the local samples to a selection from the various DVD collections of samples with commentaries we created during the early years of development. The school coordinators then meet with the district coordinators to share both local and 'official' samples, discuss any issues, and 're-calibrate' as a group, feeding back to the teachers in their own schools. At the time of development, this was the most extensive programme of professional development support for English Language school-based assessment that Chris Davison and I were aware of.

3.2.1. Statistical Moderation

Added to this school-based approach to accountability, or what we might call *social moderation*, there is also a moderation of results across schools by the HKEAA. The process of statistical moderation was added by the HKEAA during the introductory year of the SBA and continues to the present. The consultancy team argued against this added process, but it was felt by HKEAA and by the advisory committee that, given the negative publicity the introduction of the English Language SBA had attracted, the public wanted the "reassurance" of a form of reliability they could recognise.

The official HKEAA document explaining statistical moderations says "There are essentially two ways in which differences in marking standards may affect SBA scores. First, teachers in a given school may be either harsher or more lenient than teachers in other schools. Second, they may tend to either overly bunch students' scores together or spread them apart too much." In fact, experience with SBA English has shown that careful professional development, the processes of in-school teacher moderation, and the support of district coordinators have led to very few concerns of those kinds (Lee 2008). In statistical moderation, the mean score and the spread of SBA scores of students in a given school are compared to a 'moderator variable', which is the same students' results on the formal examination (HKEAA, 2006). In the moderation process, the internal rank order of the SBA scores remain the same and no student's marks are changed; but the means and the group overall SBA profile of each school are compared with its own profile on the whole English exam, and with the school's position in the ranking of all schools' English performance. A school's ranking will not be changed but marks may be added or deducted during moderation to bring that school closer to its moderator. In practice this results in a kind of 'norming out' of all schools. The HKEAA statistical moderation booklet is available to all teachers, and a complete explanation of this rather complex process, with diagrams, can be found at http://www.hkeaa.edu.hk/DocLibrary/SBA/booklet_sba.pdf.

Christina Lee, then the HKEAA Subject Officer for English, reported that the moderation revealed that, not only was the SBA speaking reliable, it had proved to be more reliable than the formal speaking test, and correlated better with all the other aspects of the HKCEE English Language exam (Lee 2008). However, discussions with teachers suggest that even now, few teachers really understand how the process works, and many

of them see this external intervention as a sign that their school-based assessments are not trusted.

4. CONTINUING CHALLENGES

The literature on educational innovation and assessment reform consistently tells us that reform is difficult, and innovation often does not ‘stick’. Two issues in particular have remained problematic throughout the nearly 10 years of this initiative.

4.1 Planning time

A key issue that arose with the GI was the question of planning: should students be allowed to prepare? Should they have specific planning time in their groups, and if so, how much? This is proving to be a complex question not only for SBA but in language learning as a whole. Until fairly recently, most experts would have said that planning time and overt planning were beneficial (e.g. Skehan & Foster 1997; Yuan & Ellis, 2003), but more recently evidence is suggesting that planning is beneficial only in quite specific conditions, and that too much planning time can be counter-productive to success. Ellis (2009) found that planning has a beneficial effect on fluency, but that the effects on complexity and accuracy are less clear. Lam (this volume) found that “what can be observed in the SBA assessed interactions is often not students’ *in situ* execution of interactional competence in L2, but a ‘canned’ product of students’ execution of the competence *prior to* the assessed interaction *in L1* during pre-task planning” (INSERT PAGE NOS WHEN VOLUME COMPILED). There is much that is true in this, although in all my school observations, and on all videos, the use of Chinese occurred only occasionally in groups where students were at extremely low levels of English proficiency. However, as I watched the 500+ video clips of GIs while preparing standardising samples and training exemplars for English SBA, I saw many that were over-prepared and formulaic, showing evidence of substantial ‘planning time’. I also saw some excellent, close to authentic-seeming, GIs. I saw too, videos which included students doing their own planning before beginning a formal GI: these always seemed more lively, more interactive and on some measures ‘better’ than the formal GIs after planning. Puntel Xhafaj, Muck, and de Souza Ferraz D’Ely (2011) found, in a close review of this literature, that the factors influencing the effects of planning time on task performance are manifold and very difficult to separate out, but there are good arguments for up to 10 minutes planning/preparation time in some contexts. Nitta and Nakatsuhara’s (2014) literature review found that while “findings have varied, depending on the nature of planning, task types, and proficiency levels of learners... a general consensus by these researchers is that relatively long planning times (e.g., 10 minutes) in classroom and laboratory settings provide clear benefits to task performance in terms of fluency, but to a lesser extent to complexity and accuracy” (p. 148). The SBA Guidelines for Teachers suggest up to 10 minutes planning time before a GI.

What has happened, however, is that the culture of ‘test prep’ that is ubiquitous in Hong Kong has been brought to bear on this innovation, like every other in Hong Kong. Many Hong Kong teachers have fallen prey to the influence of textbooks ‘preparing’ students

for SBA and to the influence of the cram school culture which infiltrates their classrooms with their pupils. There is no longer much that is ‘unplanned’ about the GIs ~~now~~ due to the prevalence of training materials and sets of “advice” to students. We were, perhaps, over-optimistic in the early days about the possibility of achieving ‘authentic’ discourse with the GI, given the reality of the summative role of SBA as part of the Hong Kong formal exam structure.

4.2. Fairness

Many issues deserve much more attention than I can give them here; however, the concern about the fairness of school-based assessment when used in a high stakes context such as HKCEE or HKDSE has been pervasive, and still continues, and it must be addressed. As Qian (2008) has pointed out, fairness is as important an issue for assessment for learning and school-based assessment as for traditional tests. This issue was taken very seriously in developing the Hong Kong English Language SBA, and was the focus of one of the inaugural seminars held by the consultancy team (Hamp-Lyons 2006). Qian (op cit) was very doubtful about SBA at its introduction, commenting that “fairness cannot be reasonably predicted from the SBA, at least during the initial years of its implementation”; and that “it will be technically impractical to fairly and accurately equate all assessment results from different teachers and different schools, who may apply the same set of criteria according to their own understanding, own value systems and their individualized contexts” (p. 105). These concerns were firmly in the minds of the consultancy group while planning the introduction of SBA: indeed, it was not unreasonable that traditional language testers such as Qian should have had these concerns at the early stages. Fortunately, this concern had been addressed by the authors of the ROPES review, which was approved by the Hong Kong Legislative Council. This gave the HKEAA the credibility at senior levels to support us as we went through the multiple stages of the processes of collaboration, sharing and mentoring described above. This has given the majority of teachers – those who take the SBA task as seriously as the rest of their teaching -- a common understanding of the standards expected at each level of the assessment domains. With the confidence of clear and shared standards, many teachers have been convinced that fairness can be achieved by alternative processes that are trustworthy.

In its 2013 Handbook for students, the HKEAA describes its approach to fairness:

How will fairness be ensured in SBA?

The HKEAA will:

- provide detailed guidelines, assessment criteria and exemplars to ensure consistency in teachers’ assessments;
- provide professional development training to help teachers become familiar with how to conduct the SBA of their subject(s);
- appoint district coordinators to support schools in the conduct of SBA for individual subjects;
- moderate SBA marks submitted by different schools to iron out possible differences among schools in marking standards.

[http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/SBA_pamphlet_E_web.pdf]

The last of these has been discussed above. The first three were developed and built into the support structure by our consultancy team for the SBA English Language, and over time have gradually been implemented in SBA for other subjects, with varying degrees of enthusiasm. The introduction of district coordinators grew directly out of our experience with our core group of teachers in the development year (2005); many of these enthusiastic early adopters (Hall & Hord 2006) proved instrumental in initiating change in their own schools, and went on to become district coordinators, and we did not want to lose their expertise and commitment.

Behind all these structures is the concern for fairness: but the understanding of fairness that we have tried to convey within SBA English Language is a little different from that usually understood in the Hong Kong exam culture. In Hong Kong, fairness is synonymous with reliability. We promote a view of ‘fairness’ as ensuring that every student has the opportunity to develop their knowledge and ability to the best they are capable of. That means that both learning and assessment situations should be structured to make that possible. The SBA professional development materials emphasize that, like teaching tasks, assessment tasks should be adjusted to fit the level of the individual learner: in the same class some students may read an easier text, or be asked a linguistically-simpler question, than others; but each learner should have the opportunity to show what they can do, and be stretched to do a little more. In that way, the class teacher can use her knowledge of her own students to make sure that the scores each student receives on assessable tasks are valid representations of their language at that stage of learning. Since every student is assessed on the same scales and criteria, they can be scored according to their actual performance. This approach has led to far fewer ‘no shows’ for the speaking exam as students gain more confidence in speaking; and there are far fewer students who receive zero scores because they simply cannot say anything in English.

Writing in 2014, Qian had moderated his view of SBA somewhat; reporting a small-scale study of ‘front-line’ teachers who were users of SBA, Qian found that 73% of his 33 respondents strongly supported or supported the implementation of SBA while only 12% expressed negative views. This is a much better result than in the early years of the longitudinal study of the implementation of English Language SBA. Qian concludes: “as the English SBA is an assessment-for-learning component within a traditional assessment-of-learning examination of English Language, this new component should be viewed with a different set of expectations from what is expected of a traditional examination, in terms of validity, reliability and fairness” (Qian 2014: 18). This gradual shift in attitude toward SBA and other forms of assessment for learning from an influential early opponent suggests that SBA is succeeding, if slowly. Indeed, the greatest indicator of success of the innovation is Qian’s comment that: “the English SBA is here to stay as an oral assessment component within the new HKDSE Examination of the English Language. Therefore, it is highly advisable to raise the awareness of teachers, students as well as school administrators with respect to the important role the English SBA can play in students’ learning of English Language” (p. 19).

Conclusion

From the distance of some ten years since the beginning of this assessment reform project, it seems to me that two elements have been key to the degree of success this assessment innovation has achieved: a rigorous professional development programme that continued over 6 years, and was accessible to every teacher across Hong Kong teaching students at this level; and a carefully developed and validated set of assessment criteria and standards, which were available for use in all schools. These two elements were possible because of the coming together of Chris Davison, a very experienced language teacher educator with a real interest in language assessment, and myself, very experienced in developing and implementing performance assessments and with a real interest in teacher education. I would like to think that we have moved some way towards assessment reform: but if we are to change language assessment practice we must impact teachers', and teacher educators', core beliefs and understanding about testing/assessment.

Acknowledgements

The SBA Consultancy Team were based in the Faculty of Education, the University of Hong Kong, from 2005-2012. The projects were led by Dr. (later Prof.) Chris Davison and Prof Liz Hamp-Lyons; the Project Manager was Ms Wendy Leung, the Project Coordinator Ms Karri Lam, and the IT Officer Miss Jo Wong. Many research assistants and doctoral students helped with the project at various times, and I thank in particular Gao Manman and Xie Qin for stimulating discussions. Thanks must go to the core group of almost 50 English teachers who worked with us through several years of development, piloting, feedback and implementation, and many of whom continued on to become Area Coordinators. They deserve the most appreciation of all, because without their enthusiasm, dedication and skills this innovation could not have been successful. Earlier versions of this chapter have been presented at the IAEA Conference Brisbane, 2009; the ITC Conference, Hong Kong, 2010; and the Language Testing Forum, Lancaster UK, 2011. In reflecting on and drawing together my experiences working in this important project over so many years, I acknowledge the support of the HKEAA and the Hong Kong Research Grants Council (RGC HKU 7268/04H) for funding the several rounds of this research, and the extensive assistance with the study of the SBA Consultancy Team. I am particularly appreciative of the stimulating and challenging working relationship I shared with my co-researcher and co-investigator, Chris Davison. Thanks and appreciation also go to my colleague at the University of Hong Kong, Prof. Stephen Andrews, with whom I worked on RGC Grant HKU7483/06H on assessment innovation, and the doctoral student participating in that project, Ms Yu Ying.

REFERENCES

- Andrews, S., & Fullilove, J. (1994). Assessing spoken English in public examinations- why and how? In J. Boyle & P. Falvey (Eds.), *English language testing in Hong Kong* (pp. 57-86). Hong Kong: Chinese University Press.

- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University Press.
- Black, P & D. Wiliam. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-68.
- Boyle, J. & P. Falvey (Eds.). (2004). *English language testing in Hong Kong*. Hong Kong: Chinese University Press.
- ~~Bray, M. (2012). *Wolves lurking in the shadows of education*. *South China Morning Post*, July 19, 2012.~~
- Bray, M. (2007). *The shadow education system: Private tutoring and its implications for planners*. (Fundamentals of Educational Planning 61, 2nd Ed.). Paris: UNESCO International Institute for Educational Planning.
- ~~Bray, M. (2012). *Wolves lurking in the shadows of education*. *South China Morning Post*, July 19, 2012.~~
- Carless, D. (2011). *From testing to productive student learning*. New York: Routledge.
- Chan & M. Bray. (2014). Marketized private tutoring as a supplement to regular schooling: Liberal Studies and the shadow sector in Hong Kong secondary education. *Journal of Curriculum Studies*, 46 (3), 361-388.
- Chen, Q. & Klenowski, V. (2009). Assessment and curriculum reform in China: the College English test and tertiary English as a foreign language education. In: *Proceedings of the 2008 AARE International Education Conference*, 30 November - 4 December 2008, Queensland University of Technology, Brisbane.
- Cheng, L. (2011). Supporting student learning: Assessment of learning and assessment for learning. In D. Tsagari & I. Csepes (Eds.), *Classroom-Based Language Assessment*. Language Testing and Evaluation Series (Volume 25) (pp.191-203). Frankfurt: Peter Lang.
- Cheung, D. 2001. School-based assessment in public examination: Identifying the concerns of teachers. *Educational Journal* 29 (2), 105-123.
- Choi, C-c. & C. Lee. (2009). Developments of English Language assessment in public examinations in Hong Kong. In L. Cheng, & A. Curtis (Eds.) *English language assessment and the Chinese learner* (pp. 60-76).New York: Routledge.
- Curriculum Development Council (1999). *English Language teaching syllabus and the Senior Secondary curriculum*. Hong Kong: Hong Kong Education Bureau.
- Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly*, 4(1), 37-68.
- Davison, C. (2013). Innovation in assessment: Common misconceptions and problems. In K. Hyland & L. L. C. Wong (Eds.), *Innovation and change in English language education*. (pp. 263-275). Abingdon: Routledge.
- Davison, C. & Hamp-Lyons, L. (2010). The Hong Kong Certificate of Education: School-based Assessment Reform in Hong Kong English Language Education. In L. Cheng and A. Curtis (Eds.). *English language assessment and the Chinese learner* (pp. 248-262).New York: Routledge.
- Education Commission of Hong Kong (2004). *Learning for life: Progress Report on Education Reform* (3). Hong Kong: HKSAR Education Commission. Download from: http://www.e-c.edu.hk/eng/reform/progress/progress003_full_eng.pdf
- Education Commission of Hong Kong (1997). *Quality School Education*. Hong Kong: HKSAR Education Commission.

- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics* 30 (4), 474–509.
- Elman, B. (2000). *A cultural history of civil examinations in late imperial China*. Los Angeles: University of California Press.
- Evans, S. (2000). Hong Kong's new English language policy in education. *World Englishes*, 19 (2), 184-204.
- Fok, W-k. (2012). HKCEE English Language school-based assessment: Its implementation at the frontline, University of Durham EdD thesis, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/3522/>
- Fok, P-k, Kennedy, K., Chan, K-s J.. & Yu, W-m F. (2006). Integrating assessment of learning and assessment for learning in Hong Kong public examinations: Rationales and realities of introducing school-based assessment. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore, 21-26 May 2006. Accessed on 14 January 2013. http://www.iaea.info/documents/paper_1162a1b7ea.pdf
- Gan, Z-d, Davison, C. & Hamp-Lyons, L. (2009), Topic Negotiation in Peer Group Oral Assessment Situations: A Conversation Analytic Approach. *Applied Linguistics* 30, 3: 315–334.
- Gao, M-m (2011). School-based assessment in Hong Kong: The impact on students' attitudes and anxiety. Hong Kong University. PhD dissertation. Available at HKU Scholar's Hub: <http://hub.hku.hk/handle/10722/134089>
- Hall, G. E., & Hord, S. M. (2006). *Implementing change: Patterns, principles, and potholes*. Boston, ME: Allyn and Bacon.
- Hamp-Lyons, L. (2006). Fairness as an issue in school-based assessment. Inaugural Seminar Series on English Language School-based Assessment: Integrating Theory and Practice, Jan 9th 2006. Hong Kong University.
- Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In Cummins, J. & Davison, C. (Eds). *The International Handbook of English language teaching, Vol. 1*. (pp. 487-504). Norwell, MA: Springer.
- Hamp-Lyons, L. (2009). Principles for large-scale classroom-based teacher assessment of English learners' language: An initial framework from school-based assessment in Hong Kong. *TESOL Quarterly*, 43 (3), 524–530.
- Henrichsen, L. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956–1968*. Westport: Greenwood Press.
- Hong Kong Legislative Council (Legco). (2003). LC Paper No. CB(2)634/03-04(01). Downloaded 12 February 2009 from <http://www.legco.gov.hk/yr03-04/english/panels/ed/papers/ed1215cb2-634-1-e.pdf>
- IBM. (2003). *Strategic review of Hong Kong Examinations and Assessment Authority: Final Report*. Examinations Authority: Hong Kong. Downloaded 10 Sep 2007 from: http://www.hkeaa.edu.hk/doc/isd/Strategic_Review.pdf
- Kellaghan, T., & Greaney, V. (1992). *Using examinations to improve education: A study in fourteen African countries*. Washington, DC: World Bank.
- Kennedy, K. (2013). High stakes School Based Assessment and cultural values: Beyond issues of validity. Key Note Address, Seminar on 'School based assessment: Prospects and realities in Asian contexts', 3 June 2013, Kuala Lumpur, Malaysia.

Downloaded 11 November 2014 from

<http://www.cambridgeassessment.org.uk/Images/139719-sba-seminar-papers.pdf>

- King, R. (1994). Historical survey of English language testing in Hong Kong. In J. Boyle & P. Falvey (Eds.), *English Language Testing in Hong Kong* (pp. 3-30), Hong Kong: Hong Kong Chinese University Press.
- Lee, C. (2008). The beneficial washback of the introduction of a School-based Assessment component on the Speaking performance of students. Paper presented at the 34th IAEA Conference, Cambridge, September 2008. Downloaded 13 August 2014 from http://iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/152128_Lee.pdf
- Lee, I. (2005). Latest initiative adds to burden. *South China Morning Press*, May 7 2005. Downloaded 1 April 2008 from: <http://www.scmp.com/article/499634/latest-initiative-adds-burden>
- Nitta, R. & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing* 31(2), 147–175.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(1), 109-148.
- Puntel Xhafaj, D.C., Muck, K. E. & de Souza Ferraz D'Ely, R. C.. (2011). The impact of individual and peer planning on the oral performance of advanced learners of English as a foreign language. *Linguagem & Ensino, Pelotas*, 14 (1), 39-65 (Jan./Jun. 2011).
- Qian, D.D. (2008). English language testing in Hong Kong: A survey of practices, developments and issues. *Language Testing*, 25; 85-110.
- Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice*, 21 (3), 251-270.
- Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1 (3), 185-211.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory* (Studies in Language Testing Series, Volume 22). Cambridge: Cambridge ESOL and Cambridge University Press.
- Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24 (1), 1-27.
- Yung, B. H. W. (2002). Same assessment, different practice: Professional consciousness as a determinant of teachers' practice in a school-based assessment scheme. *Assessment in Education: Principles, Policy & Practice*, 9 (1), 97-117.