

Article

English Language Assessment in Hong Kong: A 50-year Retrospective

David Coniam

The Education University of Hong Kong, China

Abstract

This paper presents a personal picture of my long-standing association with the English language teaching and assessment situation in Hong Kong. The paper presents a 50-year retrospective of assessment in Hong Kong, through the lens of my 40 years in the territory and my personal experience of English language teaching, teacher education, and assessment. I present a historical and theoretical picture of how English language examinations have moved forward in Hong Kong, and how I was fortunate enough to be involved in the big changes which were taking place in assessment and English language assessment in particular. While the picture I portray through this paper is a somewhat personal one, it contributes to an understanding of how assessment reform has been forward-looking, and largely successful, in Hong Kong, painting a picture of how assessment development has complemented curriculum development. I suggest that it may be instructive for educators in other jurisdictions to consider the long-term picture of development in English language assessment reform in their own country with a view to analyzing their own perspectives of the relative success of policy changes and large-scale reforms.

Keywords

Assessment, English language, Hong Kong, retrospective

1 Introduction

This paper details not only my 40-year sojourn in Hong Kong and my long-standing association with its English language assessment situation but also a 50-year journey through assessment in Hong Kong, augmented by my personal experience of years of English language teaching, teacher education, and assessment. I present a historical and theoretical portrait of how English language examinations have moved in Hong Kong – onward and upwards, when I was fortunate enough to be in the middle when big changes were taking place in English language assessment: from how assessment was conceptualised to how it was delivered.

Address: Department of Curriculum and Instruction, the Education University of Hong Kong,
Block D1-1/F-38, 10 Lo Ping Road, Tai Po, Hong Kong, China

E-mail: coniam@eduhk.hk

The picture I illustrate is a personal perspective, based on my experiences and perceptions of the Hong Kong situation. The major issues discussed, however, will reflect development in English language teaching and assessment that many jurisdictions in Asia have been grappling with over the past half century in terms of different types of curriculum and assessment reform and the extent to which such reforms have been embraced. Pictures of curriculum reform have been published for a number of Asian countries. Ho (2002) presented snapshots of different countries in East Asia, while specific country analyses were reported by Boyle (2004) on Hong Kong, Wang and Lam (2009) on China, and Choi (2015) on Japan and South Korea. The current paper adds to our repository of knowledge, complementing the picture of curriculum development, by providing a specific blueprint of assessment development in one jurisdiction.

I first present a brief picture of the Hong Kong education and examination system, to put the educational context – and the place of English language teaching in that context – into perspective. To provide an anchoring backdrop of my experience, I frame issues within the context of the key test quality concepts *Validity*, *Reliability*, and *Washback* – borrowing in part from the work of Li (1997). The paper then moves through assessment in Hong Kong one decade at a time, with my experiences framed as appropriate against a relevant key test quality concept. The paper closes by making reference to how such reflection may be conducted profitably in a broader Asian context, with a view to gauging development in different countries and regions.

The backbone of English language assessment in Hong Kong is the public examinations body, the Hong Kong Examinations (and Assessment) Authority (HKEAA). The HKEAA was established in 1977, prior to which, public exams had been administered under the aegis of the then Education Department (ED) (see Choi & Lee, 2010, p.60). I have had (and still maintain) a long-standing association with the HKEAA (having spent my formative years there in the late 1980s), and in fact still do. Consequently, a considerable amount of my presentation in this article relates to my experience and association with the HKEAA. A recent plenary presentation I made for the HKEAA (Coniam, 2018) centred around my half a lifetime of engagement with public examinations and public examination data.

2 The Hong Kong Education and Examination Systems

Hong Kong has, since 2009, provided twelve years of compulsory education. Formal education, which begins at age six when children enter primary school, extends six years (from Primary 1 [P1] to Primary 6 [P6]). Chinese (standard written Chinese / spoken Cantonese) is the medium of instruction in most primary schools, with the primary curriculum covering a wide spectrum of subject areas including Social Studies, Science, Chinese, English, Mathematics, Music, Arts and Physical Education. The allocation of students to secondary school is based upon students' examination results in P5 and P6.

Secondary school education lasts six years (from Secondary 1 [S1] to Secondary 6 [S6]). Secondary schools are streamed ('banded') into three broad bands of ability, each band covering approximately 33% of the student ability range. In three quarters of public secondary schools (of which there are approximately 400), the medium of instruction is Cantonese; in the other quarter, the medium of instruction is English.

Prior to 2012, secondary schools in Hong Kong operated on a British "5+2" model, with two major public examinations. The Hong Kong Certificate of Education Examination (HKCEE) was administered at the end of Secondary 5 (Year 11). After Secondary 5 (Year 11), students could, in principle, continue in full time education – although there were only places for approximately 40% of the Year 11 cohort to continue on to Year 12 studies. At the end of Secondary 7 (Year 13), students then sat the Hong Kong Advanced Level Examination (HKALE) in three or four subjects. The HKALE results were also used for university entrance purposes. Since 2012, there has been only one major public examination in Hong Kong – the Hong Kong Diploma in Secondary Education (HKDSE), which is held at the end of

Secondary 6 (Year 12). In addition to four core subjects – English, Mathematics, Chinese and Liberal Studies – students choose a further two or three elective subjects. The ‘loss’ of one year in the school system was mitigated by the addition of one further year in undergraduate education extending it from three to four years.

Formal English language classes are offered from P1. Standard provision of English language in Hong Kong primary and secondary schools (across the whole 12-year period of compulsory education) is approximately 4-5 hours per week for the duration of the academic year.

3 Test Quality Concepts

One of the major issues that needs consideration at the outset is: what is the purpose of the English language curriculum, and how does this translate into curriculum and assessment aims? Half a century ago, the purpose of the English language curriculum might have been seen as “supplying skilled manpower for trading, communication and commercial purposes”, with direct curriculum ‘aims’ expressed as:

1. Mastering every grammatical structure, and
2. Testing what students knew about English grammar.

Such a limited focus meant that the purpose of the English language curriculum did not really achieve its goals (see e.g., Nyland, 1990).

Through major curriculum development from the late 1970s onward with the advent of a Communicative Approach to ELT (e.g., Littlewood, 1981), the focus shifted from a single focus on structure to one involving a more cognitive, affective humanistic approach, with ‘communication’ being a goal as important as structure (e.g., Richards & Rodgers, 2001).

The current ‘aims’ of the English language curriculum may now therefore be framed as:

1. Students communicating in English, and
2. Testing what students can do in English.

The key concept here is *Validity*. Validity (see e.g., Bachman & Palmer, 1996; Messick, 1989) may be framed as constituting (1) the ‘skills’, ‘abilities’, and ‘constructs’ that are being tapped in the test, and (2) the extent to which a given test score can be interpreted as an indicator of the abilities or constructs to be measured.

The second key concept is *Reliability* (see e.g., Hughes, 2003), which relates to how results awarded to test-takers change across periods of time, across different groups of students, and between markers. Key factors affecting Reliability are (1) the degree of objectivity that exists in a test; (2) test length; and (3) the amount of question choice offered to candidates.

The third key concept is *Washback* (see e.g., Alderson, 2004; Cheng, 2005; Choi & Lee, 2010), which relates to the effect that changes to examinations have on teaching. Despite the fact that the HKEAA is an independent examinations body, positive washback has been at the forefront of many of the major changes that have occurred with its English language examinations, and it has taken very seriously the notion that examinations should encourage worthwhile classroom practices.

Following each analysis, drawing as mentioned on the ideas of Li (1997), I present a synthesis of the three concepts (Validity, Reliability, and Washback) as a ‘VRW triangle’, the concept here being that the (red) dot in the centre gets pulled in the direction of which concept(s) has most effect on it in a given analysis. Figure 1 presents the ‘triangle’, and its default position.

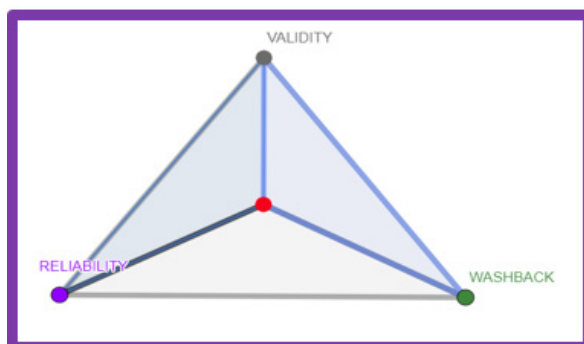


Figure 1. VRW Triangle

In order to make the material more digestible, and to put issues into perspective, I will frame issues as I move through the paper through the lens of decades. Although I was not in Hong Kong in the 1960s and 1970s, I have gained access to past ED/HKEAA documentation – which helps to fill in the gaps.

4 The 1960s and 70s – Behaviourist Times

The 1960s and 70s were Behaviourist times. In line with Behaviourist principles there was a strong focus on reliability, and accuracy was the order of the day (see Howatt, 2004). In tandem with the ‘methodology’ underpinning Behaviourist principles, the activities that predominated were translation, grammar drills, and a substantial amount of multiple choice questions. I will illustrate with a couple of examples.

One section of the 1967 School Certificate Examination, English Paper III required candidates to translate from English into Chinese. Figure 2 shows a sample.

SECTION B (30 marks) Translate the following passage into Chinese

On a splendid September day, I left my native land. After a very interesting journey I arrived in London. I was amazed at the difference between my small village and the huge city. What traffic! What an uproar in the streets! At first the noise nearly deafened me, but after an hour or so I became used to it. Everything was new and strange and I must confess that my first impressions were not very favourable. When I arrived at the hotel which had been chosen for me by a friend, I felt very tired. I had seen a great deal in one day, and I felt in need of rest in mind and body that evening.

Figure 2. Passage for Translation (extracted from the 1967 School Certificate Examination)

While the text is clearly dated, and there are (from a current perspective) some non-politically-correct elements (‘my native land’), there are a number of points worth considering.

It is not a young person’s text. It is written by a middle-aged examiner (‘I must confess that my first impressions’; ‘I felt in need of rest in mind and body’). This is not how a young person speaks – or spoke 50 years ago. Times have changed and the genre and makeup of texts presented to 16-year-old students are now more relevant to them (see, for example, Krashen & Terrell, 1983 on the ‘Natural Approach’).

Testing points have been chosen to assess a range of elements: past tense, passive, ‘What a ...’, relative clauses. It is not a spoken text, although it tries to appear that way. There are complex sentences and embedded relative clauses.

Another feature of past English language examinations was a clear focus on grammar. The sample in Figure 3 below – from the 1966 Secondary School Entrance Examination (SSEE) – required candidates to transpose sentences from indirect speech into direct speech. This is a fascinating exercise in that it is one that would almost never take place in regular communication, either written or spoken. Validity in this task is, consequently, very low.

- a. The old man said that he had seen the doctor.
 a. “ _____ ”, the old man said.
- b. The teacher said that we were being very naughty.
 b. “ _____ ”, said the teacher.
- c. She asked me when he would arrive.
 c. “ _____ ”, she asked me.
- d. He told John to open the door.
 d. “ _____ ”, he said to John.
- e. We asked the girl what her name was.
 e. “ _____ ”, we asked the girl.

Figure 3. Indirect Speech into Direct Speech Exercise (extracted from the 1966 SSEE)

In line with a focus on reliability, there was a strong emphasis on multiple-choice testing – that first appeared in Hong Kong English language examinations in 1969 (King, 1994).

In order to eliminate the possibility of candidates cheating, multiple forms of the same test were created, with different lettering for the options (ABCD, EFGH, WXYZ), reordered options, and the key placed in a different place. Figure 4 presents a (mocked-up) sample.

1. The _____ of sleepless nights made her ill.
 A. strain [KEY]
 B. stress
 C. tension
 D. pressure
4. The _____ of sleepless nights made her ill.
 Q. stress
 R. strain [KEY]
 S. tension
 T. pressure
9. The _____ of sleepless nights made her ill.
 W. strain [KEY]
 X. stress
 Y. tension
 Z. pressure

Figure 4. Mockup of Multiple-form MC Test

A major issue with assessment in the 1960s and 70s was the effect on teaching caused by the format of the examinations (MC in particular) and negative washback (see Alderson & Hamp-Lyons, 1996).

While MC testing was an accepted part of the culture, the HKEAA was concerned about its negative washback, and strove to minimise the fallout by trying to restrict the amount of MC practice paper work that schools might do for the English language public examinations. They performed this by not publishing the MC papers from the examinations (King, 1994). Unfortunately, this did not prevent teachers from getting hold of the papers since a classic workaround was for a teacher to ask each student in their class to memorize two or three MC questions and to write them down for the teacher immediately after the examination. In this light, Figure 5 presents the VRW interpretation, with the right-hand box indicating how the dot, the apex, has shifted.

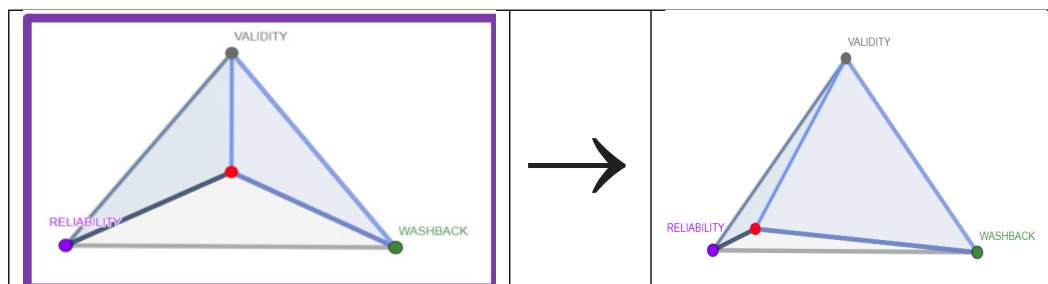


Figure 5. VRW Triangle

The three examples above pull the triangle's apex very strongly towards Reliability, at the expense of Validity, and Washback.

In the 1970s, while there was a continued focus on reliability, meaning and relevance were beginning to enter English language examinations. The Year 13 Hong Kong Advanced Level Examination Use of English (UE) examination – The University of Hong Kong's (HKU) entrance test – was very reliability-focused. It did nonetheless include test types and material which were 'relevant' to tertiary-level studies: sections of the examination included cursory reading (albeit multiple-choice) and an academic lecture listening test.

5 Use of English Listening and Oral Tests: Format and Teaching Approach

It is worth dwelling awhile on the format of the test and how this affected the way that teachers approached the teaching of listening in the 1970s in the light of what their students were to face in the Use of English Listening Test. The format in which the test was delivered was less natural than the current format where candidates have time to look over the question booklet before hearing the listening input. They have the opportunity to focus on what the test will be about before they hear (usually once) the tape script. In those days, the listening test was played twice; the candidates listened 'blind' in that they did not receive the question booklet until after having listened twice. The intended objective was that – as with an academic lecture – they would take notes, and later make sense of them to answer the questions.

Unfortunately, this objective (and test validity along with it) was, however, widely circumvented by the 'two-pen approach' devised by smart (test-wise) Hong Kong teachers. Under this 'two-pen approach', candidates took a blank sheet of paper, which they divided in half vertically with a line down the middle. During the first listening, they made notes down the left-hand side in blue. On the second listening, they took a different coloured pen (e.g., red) and made notes down the right-hand side of the page in red. Finally, they opened the question booklet, and attempted to patch their notes together to answer the questions (see King, 1994). The activity was more like speed dictation than a listening test where meaning was being assessed in the context of a stream of speech. Consequently, the Listening Test in this format was low in validity and was one reason for the subsequent overhaul of the Use of English examination in 1989 (see below).

The oral test component of the Hong Kong Certificate of Education Examination (HKCEE) English comprised three parts: (1) *Reading aloud a dialogue*; (2) *Discussing a picture*; (3) *Having a conversation with the examiner(s)*. The oral test was introduced in 1974 and ran in this format until 1995. The oral test was reliability focused, with an emphasis on accuracy rather than fluency (see Coniam, 1990). This was evident in the first part *Reading aloud a dialogue*, where the candidate and examiner read a dialogue together. The more open Part 3 – *Conversation with the examiner(s)* – in principle allowed for fluency work, but in practice this was more akin to an ‘inquisition’ in the manner in which two examiners (the two people behind the desk in the figure below) tended to ‘interrogate’ a candidate, as the graphic in Figure 6 below is meant to suggest.



Figure 6. Oral Exam ‘Inquisition’

Nonetheless, while the ‘conversation’ was more like an inquisition, it was a direct test with the entire candidature, and it did require candidates to speak a bit of English. This was quite an achievement in the very behaviourist-oriented education system prevailing in Hong Kong in the 1970. Consequently, I would frame the VRW triangle as in Figure 7, where the apex has moved away from its strong attachment to Reliability alone.

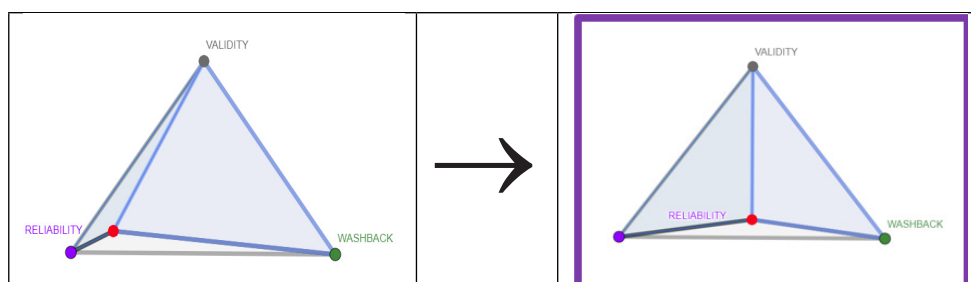


Figure 7. VRW Triangle: HKCEE Oral Test 1975-1995

6. The 1980s – A Communicative Approach to Language Teaching

In the context of a worldwide movement that advocated that there was more to language teaching than merely grammar (Hymes, 1972), the 1980s saw the advent of a Communicative Approach to Language Teaching.

In line with the principles of a ‘Communicative Approach’ – and the needs of society/business expanding – elements other than grammar began to come into focus in both school curriculums and English language examinations. There began to be a greater focus on language use, which in examination terms meant greater validity, in that the examination score gave more of an indication of what candidates could **do** in English than previous examination formats did (Messick, 1989).

The effect of the new communicative movement was major revisions to the key HKCEE and HKALE (see Choi & Lee, 2010). Multiple-choice and grammar testing were still part of the public examinations, but were being quietly de-emphasised.

One major innovation to the HKCEE of English language – and which involved a major commitment from the HKEAA – was the introduction, in 1986, of a listening test. The HKCEE Listening Test was more general in its orientation than the Year 13 Use of English Listening Test depicted above, which, as The University of Hong Kong's entry test, had the format of an academic lecture. Since the radio signal was not strong enough to cover the whole of Hong Kong, the listening test involved the school halls of most secondary schools being equipped with 'induction loops'. Nevertheless, the listening test still required five parallel sessions, and the HKEAA had 25,000 sets of headphones for each session – a logistical nightmare (see King, 1994).

In 1989, the Use of English exam was completely revised (see King, 1994). It was no longer solely HKU's entrance test, but was intended to perform the dual role of an entrance test for all tertiary institutions and also be a valid assessment for Year 13 school leavers who would join the workplace, working for a company or business.

The 1989 revision of the Use of English examination was therefore much more communicative in its orientation – with better validity; there was still a reliability focus, however. Nonetheless a major focus of the HKEAA's was the effect of washback: having students do things in their English language classrooms that would have wider relevance than merely a university entrance test. One major omission in the washback picture, though, was the fact that there was still no oral test in the Use of English – a situation that was not rectified until 1994.

At this point, thus, we can make a major statement. It is that the effect of the examination on teaching *did* matter in the 1980s. It was a major concern of the HKEAA's. It would be interesting to compare this 'imperative' in the context of other Asian nations or jurisdictions in the 1980s, some of whom are still trying to incorporate some elements of a communicative approach to teaching and testing in their school English language curricula and examinations even now, in the 2010s (see Choi, 2015, for a comparative discussion of the cases of Japan and South Korea).

7 The 1990s – Focus on Language Use Extended and Focus on Standards

7.1 Focus on language use

Following greater adoption of the principles of a 'Communicative Approach', language use began to come more to the fore. As a consequence, the focus on grammar, as well as on multiple-choice, was less emphasised.

In 1996, the HKCEE of English language underwent radical revision. The previous listening test, which had all been multiple-choice, was incorporated into an 'integrated' listening/reading/ writing paper. Cheng (2005) investigated this change from the perspective of washback and reported how the modification of an examination changes teachers' classroom practice. She stated that changes in teaching content were the most obvious indicators, with observed changes to classroom activities in line with the new more communicative examination (Cheng, 1997, p.49).

The oral component of the examination was radically revised as well. This moved from its previous 'interrogation' format to one where the major part was a group discussion. Previously the examiners had dominated the discussion; in the revised version of the oral the examiners were only assessors, taking no part in the discussion at all. Figure 8 presents a sample. (The photo in Figure 8 was actually taken during the SARS outbreak in 2003 – see below).



Figure 8. Group Oral – Configuration (taken during 2003 SARS outbreak)

In this mode of delivery of the oral, for which Figure 7 presents a sample, Validity was therefore enhanced. (It should be noted nonetheless that with a view to maintaining reliability; considerably more training and standardisation was required and provided.) Figure 9 illustrates how the triangle's apex has shifted from its position in Figure 7.

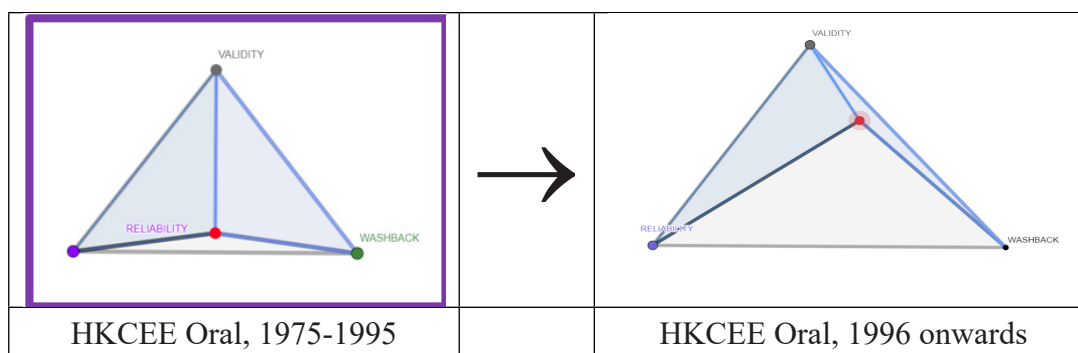


Figure 9. VRW Triangle: Shift in Oral Exams

7.2 Focus on standards

The mid-1990s also saw a strong focus on standards – teacher English language standards in particular.

A major initiative by the Government of the Hong Kong Special Administrative Region (HKSAR) involved establishing minimum language proficiency examinations (also known as ‘language benchmarks’) for all teachers in Hong Kong primary and secondary schools. The genesis of these benchmark examinations lay in concern – expressed since the early 1990s by different sectors of the business and education communities in Hong Kong – over perceived falling language standards especially after the publication of research conducted in the early 1990s that revealed that less than 20% of the secondary workforce of Hong Kong’s English language teachers were both academically and professionally qualified (Tsui et al., 1994). The government (at the behest of the independent Education Commission) therefore deemed it essential that teachers of English develop their second language skills as one of the prerequisites for being able to teach and adapt to new assessment methods and curricular objectives in their classrooms.

A major initiative by the Education Commission (1996, p. 11) requested that minimum language proficiency standards be specified for English language teachers. I will now briefly detail my involvement with the development of the ‘Language Proficiency Assessment of Teachers of English’ (LPATE), as it became known, from an initial study in 1996 to the first administration of the test in 2001.

Following an initial consultancy report in 1996, an initial test battery was constructed to assess teacher English language standards. This battery of tests comprised a set of ‘formal’ (i.e., Reading, Writing, Listening and Speaking Tests) as well as an observation of two live lessons (the Classroom Language Assessment performance test). The latter test was considered to be the most valid part of the test battery since it consisted of a performance test during a genuine *target language use* situation (see Coniam & Falvey, 1999). Figure 10 below depicts the apex of the triangle in the light of the Classroom Language Assessment (CLA) test, with its criterion-referenced scales and accompanying descriptors.

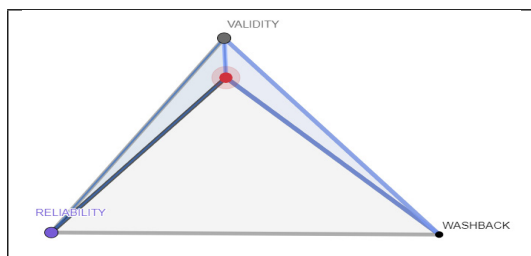


Figure 10. VRW Triangle for the LPATE's Classroom Language Assessment

Following some initial trialling and refinement of the tests, a pilot study – the Pilot Benchmark Assessment (English) [PBAE] – was administered in early 1999 to a representative sample of English language teachers teaching English at Years 7-9. The key issue was an investigation of how the prototype language ability tests – and in particular the CLA component and its attendant benchmark levels fit this stratum of English language teachers. Outcomes (and reactions by teacher participants taking the tests) were positive. The formal examination syllabus and specifications for the LPATE were then published in 2000 (Government of HKSAR, 2000), with the first live administration of the LPATE taking place in March 2001 (see Coniam & Falvey, 2013).

Following the publication of the examination syllabus and specifications in early 2000, two key decisions were taken by the HKSAR government. One, teachers holding a relevant degree and a professional teaching qualification were exempted from having to take the test; and two, the HKSAR Government earmarked (US\$30 million) for in-service development and immersion courses for teachers to attain the required standards.

While it is not possible to directly quantify that the LPATE has raised English language teachers' language standards, Drave (2006) suggests that the LPATE has had a positive impact in that there is now an acceptance that all English teachers must now meet the minimum standard required if they are to be permitted to teach. From a government-commissioned grant in 2015, I conducted a qualitative study that investigated the perceptions of 24 long-serving teachers of their impressions of the impact of the LPATE over the 15 years since its inception. The LPATE was, in the main, perceived as having had a positive impact. The perception was that English language teachers' language standards – along with their language knowledge and pedagogical skills – had developed as a result of the test and through formal training, over the 15-year period of the administration of the LPATE (see Coniam et al., 2019).

8 The 2000s – Restructuring of Secondary Education; Onscreen Marking; SARS

Returning to the chronological timeline, three issues dominated the 2000s. The two that stood out were major changes to both the Hong Kong education and examination systems. A third issue was the government and public reaction to Severe Acute Respiratory Syndrome (SARS).

As mentioned earlier, Hong Kong's education system underwent significant restructuring in 2009. Under the restructuring, secondary education now lasts six years with a single public examination (the

Hong Kong Diploma in Secondary Education [HKDSE]) administered at the end of Year 12 (age 18). In line with the drastic structural changes to the whole education system about to take place in 2009, the examinations themselves (of which English language was at the forefront) also saw massive changes to examination content and format, to marking, and to grading in 2007. Onscreen marking (OSM), which I will discuss below, was another key innovation which began to come in in the 2000s.

8.1 2007 revised HKCEE English language exam

The revisions of the HKCEE English language examination in 2007 presaged major changes that were to be implemented for all subjects with the advent of the HKDSE in 2012. This was the biggest ‘upheaval’ ever for English language examinations, and a number of major changes were implemented. In an interesting adjustment of policy, there was a much greater rapprochement between the HKEAA and the Curriculum Development Institute (CDI) than in the past. Effectively the new syllabus was produced by the HKEAA in conjunction with the CDI.

The English language examination became standards-referenced (as opposed to the strict norm-referencing that had long dominated the Hong Kong examination system), school based assessment was introduced (see Davison, 2007; Davison & Hamp-Lyons, 2009). There were also significant changes to the format of the English language examination, whereby on each examination paper, a single theme (schema) ran through the paper, rather than as previously when a paper consisted of a set of unrelated subtests.

From a number of perspectives, the validity of the English language examination was enhanced. The Assessment For Learning aspect of School Based Assessment (SBA) made it possible for students to relate more to their peers and the material in the examination than having to sit an examination with a bunch of strangers in an examination hall.

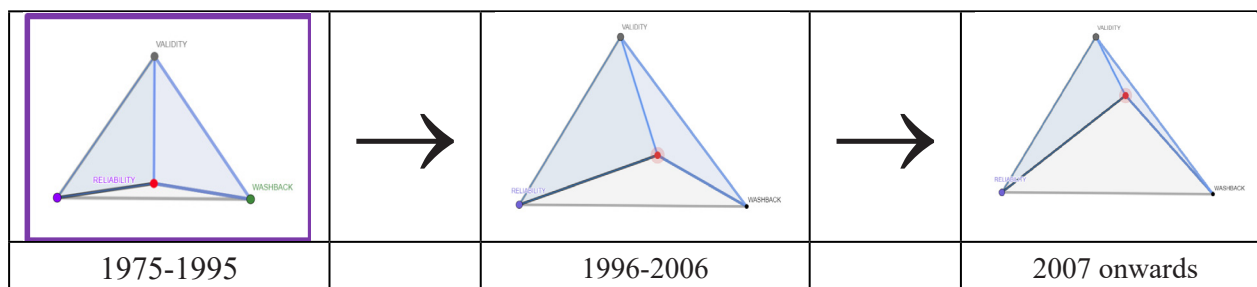


Figure 10. VRW Triangle: Shifts in VRW of HKCEE Oral Exams

Given the long history of norm-referencing, and the more ‘humanistic’ approach of criterion referencing, a question raised was whether standards would ‘slip’ as more students achieved potentially higher marks. This did not occur, however. HKCEE pass rates remained constant, and markers did not appear to have ‘overmarked’.

8.2 Onscreen marking (OSM)

As mentioned, another major change to the Hong Kong assessment horizon was the manner in which examinations were marked, with, initially, the new English and Chinese syllabuses being marked on screen. This was such a major change to how examinations are marked in Hong Kong, and is an area where Hong Kong was effectively leading the world. From 2009 onwards, I conducted a series of quantitative and qualitative studies on a range of issues – among which were statistical comparability, marker technological readiness, and marker reactions to OSM and to the system. These studies are documented in Coniam and Falvey (2016).

Statistically, no differences emerged among examination papers marked – whether in the medium of English or Chinese, being single or double marked, as well as and involving long- and short-answer questions (see e.g., Coniam, 2009).

Markers felt themselves to be quite competent technologically, although new markers were in general more positive than experienced ones. Overall, however, despite certain misgivings, even experienced markers were aware of the potential benefits available with OSM – rather than the new system simply inspiring difficulties and drawbacks (see e.g., Coniam, 2013).

The picture that emerged from the different OSM studies was that buy-in and acceptance by markers clearly increased with each year. In 2012, with the examinations of all 20 HKDSE subjects marked on screen, it was very important to ensure that the system was reliable. The studies I steered reveal that this was likely to be the case. It was very rewarding to have been part of this validation process for onscreen marking – in which Hong Kong is a world leader.

8.3 Severe acute respiratory syndrome (SARS)

The other major event of the 2000s, as mentioned, was Severe Acute Respiratory Syndrome (SARS), which had a considerable impact upon the workings of the Hong Kong education and assessment system. All classes in Hong Kong were suspended during most of April 2003, with all students and teachers required to wear face masks once schools reopened. While the use of face masks posed some level of discomfort to wearers, the effect from the English language oral assessment perspective (in particular the Year 11 Hong Kong Certificate of Education Examination [HKCEE] oral test held in the month of June) was that all examiners and candidates had to interact with certain facial cues removed – see Figure 8 above. As there had been some concern that the wearing of a facemask invalidated some of the assessment results, I conducted a study in March 2004 (Coniam, 2005) to investigate whether wearing a facemask intruded on the oral assessment score. In the study, the entire Year 11 cohort (N=186) of an average ability Hong Kong secondary school took a past HKCEE oral test both with and without face masks as their mock HKCEE oral examination, replicating HKEAA examination conditions and procedures as far as possible.

Contrary to expectations, test data results did **not** suggest that face masks had an effect on test-takers' oral test scores. Non-significant results emerged on t-test analyses conducted for all rating scales used – even the *Audibility* and *Comprehensibility* scales which would have been most susceptible to the effect of the face mask. Whereas the wearing of facemasks appeared to have a deleterious effect on validity, reliability was – surprisingly to many – not affected.

9 In Closing

As will be clear from my account in this paper, Hong Kong English language examinations have come a long way in 50 years – not only theoretically, but technologically and practically. In large part, I myself developed with the English language examinations that I had contact with: from a mechanistic, behavioural orientation to one which is more humanistic, more thinking, more feeling, one which is in tune with the time, with research, and with both assessment and teaching.

While the picture I have portrayed throughout this paper has been a personal one, it contributes to an understanding of how assessment reform has been quite forward-looking, and largely successful, in one jurisdiction – Hong Kong. The picture of assessment development in the paper has been intended to complement that of curriculum development. As such, it may be instructive therefore for educators in other countries and jurisdictions to consider the long-term picture of development in English language assessment reform in their own country with a view to analysing where they stand in terms of the success of policy and of being in tune with current thinking.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280-297.
- Alderson, J.C. (2004). Foreword. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. ix-xii). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford, NY: Oxford University Press.
- Boyle, J. (2004). Linguistic imperialism and the history of English language teaching in Hong Kong. In K. K. Tam & T. Weiss (Eds.), *English and globalization: Perspectives from Hong Kong and mainland China* (pp. 65-84). Hong Kong: The Chinese University of Hong Kong Press.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge, UK: Cambridge University Press.
- Choi, C. C., & Lee, C. (2010). 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 & London: Routledge.
- Choi, T. H. (2016). Glocalization of English language education: Comparison of three contexts in East Asia. In C. M. Lam & J. Park (Eds.), *Sociological and philosophical perspectives on education in the Asia-Pacific region* (pp. 147-164). Hong Kong, China: Springer.
- Coniam, D. (1990). Reading aloud speed and language proficiency. *Hongkong Papers in Linguistics and Language Teaching*, 13, 47-53.
- Coniam, D., & Falvey, P. (1999). The English language benchmarking initiative: A validation study of the Classroom Language Assessment component. *Asia Pacific Journal of Language in Education*, 2(2), 1-35.
- Coniam, D. (2005). The impact of wearing a face mask in a high-stakes oral examination: An exploratory post-SARS study in Hong Kong. *Language Assessment Quarterly*, 2(4): 235-261.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Coniam, D. (2013). The increasing acceptance of onscreen marking – the ‘tablet computer’ effect. *Journal of Educational Technology & Society*, 16(3), 119-129.
- Coniam, D. (2018). *Half a lifetime of engagement with public examinations*. Research Forum on Using Assessment Data for Better Learning and Teaching, Education Bureau Kowloon Tong Education Services Centre, 31 October 2018.
- Coniam, D., & Falvey, P. (2013). 10 years on: The Hong Kong Language Proficiency Assessment for Teachers of English (LPATE). *Language Testing*, 30(1), 147-155.
- Coniam, D., & Falvey, P. (2016). *Validating Technological Innovation: The Introduction and Implementation of Onscreen Marking in Hong Kong*. Singapore: Springer.
- Coniam, D., Falvey, P., & Xiao, Y. (2019). A qualitative inquiry into the impact of the Hong Kong LPATE over the past decade and a half: Changes, challenges, and implications. *Language Assessment Quarterly*, 16(1), 105-124.
- Davison, C. (2007) Views From the Chalkface: English Language School-Based Assessment in Hong Kong, *Language Assessment Quarterly*, 4(1), 37-68.
- Davison, C., & Hamp-Lyons, L. (2009) The Hong Kong Certificate of Education: School-based

- assessment reform in Hong Kong English language education. In L.-Y Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner*. New York: Routledge.
- Education Commission. (1996). Report Number 6. Hong Kong: Government Printer.
- Ho, W. K. (2002). English Language Teaching in East Asia Today: An Overview. *Asia Pacific Journal of Education*, 22(2), 1-22.
- Howatt, A. P. R. (2004). *A history of English teaching* (2nd ed.). Oxford, NY: Oxford University Press.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Hymes, D. (1972). On communicative competence. In J. B. Pride (Ed.), *Sociolinguistics* (pp. 269-293). Middlesex, UK: Penguin Books.
- 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. 3-29). Hong Kong: The Chinese University Press.
- Krashen, S., & Terrell, T. (1983). *The natural approach: language acquisition in the classroom*. Oxford, NY: Pergamon Press.
- Li, Xiaoju. (1997). *The science and art of language testing*. Changsha: Hunan Education Press.
- Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge, UK: Cambridge University Press.
- Nyland, K. (1990). English use in business – forming a link with English education in Hong Kong. In V. Bickley (ed.) *Language use, Language Teaching and the Curriculum*. Hong Kong: Institute of Language in Education.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Richards, J. C., & Rodgers, T. S. (1986/2001): *Approaches and methods in language teaching: A description and analysis*. Cambridge, UK: Cambridge University Press.
- Tsui, A. B. M., Coniam, D., Sengupta, S., & Wu, K.Y. (1994). Computer-mediated communication and teacher education: The case of TELENEX. In N. Bird, P. Falvey, A. B. M. Tsui & A. McNeill (Eds.), *Language and learning*. Hong Kong: Government Printer.
- Wang, W., & Lam, A. S. L. (2009). The English language curriculum for senior secondary school in China: Its evolution from 1949. *RELC Journal*, 40(1), 65-82.

David Coniam is Chair Professor and Head of Department of the Department of Curriculum and Instruction at The Education University of Hong Kong, where he is a teacher educator. His main publication and research interests are in language assessment, language teaching methodology, computer assisted language learning and academic publishing.