Opposing tensions of local and international standards for EAP writing programmes: Who are we assessing for?

Introduction

This study was undertaken at XYZ University, one of the six publicly-funded universities in Hong Kong, a region of China where English is a second or third language but is the medium of tertiary education. The vast majority of XYZ University students come from Chinese Medium of Instruction (CMI) secondary schools, and have low levels of spoken and/or written English. Recent changes in the Hong Kong education structure have meant that from the 2012-13 academic year, students come to university a year earlier than in the past, receiving one year more in tertiary education and one year less in secondary. The Hong Kong Examinations and Assessment Authority (HKEAA) introduced a new school leaving exam, the Hong Kong Diploma of Secondary Education (HKDSE), which is now largely used to decide which students will be offered university places. The 2012 English language component of this exam was benchmarked against the IELTS, resulting in a HKDSE level 3 becoming the minimum passing score. Through an indirect process of alignment, this HKDSE level was predicted to equate with an IELTS score range of 5.48-5.68, which itself is indirectly aligned by Cambridge Assessment with a CEFR level around a low B2. See Figure 1[1], an extract from the IELTS website detailing this alignment. This is also the minimum English language entry standard for XYZ University.

Insert Figure 1 here

In response to these changes, most Hong Kong universities have revisited their English Language curricula and their initial EAP courses. XYZ University set its goal for students' exit English language level at IELTS 6.5 (CEFR B2+), meaning that the EAP programme should bring students as a whole group from a notional IELTS 5.5 on entry to IELTS 6.5 on exit or from a low CEFR B2 to CEFR B2(+) or low C1. To

---

[1] Downloaded 1-4-2013 from:
http://www.ielts.org/researchers/common_european_framework.aspx

those not familiar with the CEFR scale, this may sound like a rather small step, but in fact as Figure 1 shows, the CEFR B2 performance descriptors are seen as covering a range from very high B2+ (between IELTS 6.5 and 7.0) to a mid-B1 (IELTS 5.0 falls exactly on the line between CEFR B1 and B2). This range in fact covers almost the entire Hong Kong high school graduating population likely to apply to universities.

In this paper we describe the process of developing a writing assessment to fit the re-defined needs of the students, and we discuss the problems associated with attempting at the same time to fulfill the institutional expectation that the 144-hour EAP programme would raise students' scores by almost a whole CEFR scale step. We demonstrate the problems of aligning EAP needs-based domain scales and standards with the Common European Framework of Reference. Finally, we consider the political tensions created by the use of external, even international, reference points for specific levels of writing performance for all our students.

Getting ready for change

In preparation for the new cohort of HKDSE students, who were due to arrive at the start of the 2012-13 academic year, XYZ University began a lengthy process of curriculum renewal in 2008 with the intention of a full year-long pilot with the 2011 entering cohort (see Figure 2 for a timeline of this process). There was much speculation and anticipation among tertiary educators surrounding the issue of exactly what this new breed of students, who had been through a completely revamped secondary school curriculum, would be like. Although curriculum documents and sample HKDSE exam and assessment papers had been made available by the Education Bureau and Exams Authority, there was still a sense that we were navigating into unknown territory. At this time local English Language Centres established a conference, which has since become an annual event, to share curriculum and assessment blueprints, discuss the impact of the new 3-3-4 structure and the likely changes in the student population in terms of English language proficiency. One of the plenary presentations at this inaugural conference was given by a member of the HKEAA, who attempted to shed light on the make-up of the new student body. Staff from the HKEAA also visited local universities in an

attempt to explain the new exam structure and demystify this new breed of students. Nevertheless, by the time the new Diploma of Secondary Education was introduced with the September 2012 university intake, there was little information available to help English Language Centres prepare appropriately to receive them.

Insert Figure 2 here

The economic politics of English language in Hong Kong

To fully understand the context in which curriculum planners in English-medium universities in Hong Kong are working, it is useful to have an understanding of how English language proficiency is viewed both politically and economically. Hong Kong is a small semi-autonomous territory of China with very few natural resources and a fast growing population of over 7 million. Its economy depends on the growth and international reach of its service industry, and English language skills are seen as a marketable resource. Thus, for many years, the government asked the universities to collaborate on a common English language assessment for university graduates which would enable employers to be sure that job applicants had suitable English language proficiency. Universities resisted this move because a common exit assessment would enable direct comparisons between universities with very different histories and missions and such comparisons could have political consequences.

In the absence of agreement between the universities, between 2002-2014, the Hong Kong University Grants Committee (UGC) set up and operated a system called the CEPAS (Common English Proficiency Assessment Scheme: see Appendix 1) which used IELTS as a substitute for a locally-derived common standard for English proficiency. The CEPAS encouraged final year students of UGC-funded undergraduate programmes to take the International English Language Testing System (IELTS) exam, have their scores formally reported, and receive a refund of the test fee. Thus, the IELTS became a stand-in for a local standard. The creation of the CEPAS was an indicator of how important government, and especially business, in Hong Kong consider good English language skills to be.

The use of IELTS as a standard was further embedded in the Hong Kong education system as the HKEAA had for some years, through the use of its own concordance table, linked IELTS overall scores with the grades on the former English language exam at exit from 7 years of secondary schooling, the Hong Kong Advanced Level Examination in English (HKALE). The minimum grade on the HKALE for entry to university, Grade E was linked to an IELTS score range of 5.40-6.02. This corresponds closely to the alignment of the minimum passing level for the new HKDSE with an IELTS score range of 5.48 – 5.68 (see Table 1). Note though, that students entering a Hong Kong university in and after 2012 using the HKDSE will have had one year less of English language learning at secondary school than in the past.

Insert Table 1 here

Importantly, this alignment shows that students with a notional IELTS score below 5.4 in the old system or 5.48 in the new one, should not have been admitted to a university or have been in our EAP classes. At XYZ University teachers have so far not found that this to be a true representation of the proficiency levels of many of the new students entering university in the 3-3-4 structure.

These concordances were meant to make it possible to compare students' levels of English proficiency on exit from university to their levels at entry and to an international common standard. In theory, such comparisons could be used to set targets for students' English language proficiency gains while at university. However, in reality the indirectness of the comparisons and the differences between each of the exams calls into question the validity of any such comparisons. However, with the ending of the scheme in 2014 it will be more difficult to claim any relationship between university entrants' English level at school-leaving, their progress in English within university, and their English level at university graduation. This has left a significant gap, which is only slowly being filled by embryonic collaborations between sub-sets of universities on common assessments, particularly the Diagnostic English Language Tracking Assessment (DELTA)

(Lockwood 2013). The various uncertainties and pressures of this period were primarily political, spurred by the anxieties of Hong Kong's business community to maintain, or appear to maintain, a competitive edge in the quality of its workforce, and have little to do with any university's EAP curriculum or assessment. However, this makes these pressures no less 'real'. The ongoing external demand for clarity about the levels of graduates' English is reflected in the expectations for language improvement placed on university EAP programmes, each of which has had to create their own EAP curriculum and assessments to address the needs of the new intake of post-HKDSE first-years.

New needs, new solutions

In the complex context and planned macro-level changes described above in mind, XYZ university's curriculum and assessment teams embarked on a three-year development project to i) create a curriculum that would bridge the gap between the general English language foundation established at secondary school and the more sophisticated language and skills required for academic study in an English-medium (EMI) university and ii) benchmark EAP writing standards to facilitate teachers of the new EAP programme in contextualizing and interpreting the levels on the new rubric. This process began with a large scale needs and target situation analysis to gain a clear understanding of students' capabilities at the point of entrance to EAP instruction and the expectations of what students should be able to do before they are ready for mainstream subject courses.

The 2008-2009 needs analysis in the ELC aimed to identify the contextual needs within XYZ University in order to compare them with the students' current English language capability. This included interviews with faculty members, investigation of assessment task types employed in the wider university, staff and student questionnaires, focus group interviews with students and analysis of student work.

Students' background and needs

A key finding was that most departments in the university favoured reports/essays, oral presentations and group projects as their usual assessment tasks (Lui et al., 2010). Lecturers commented that it can be difficult to decipher students' meaning due to poor language ability but that they tend to grade work for content and not language. Evans and Green (2007) in their survey of Hong Kong tertiary students found that —the subjects' problems centre on the lexical and grammatical aspects of writing‖ but comment that —an English programme that seeks to address these weaknesses by focusing on remediation or general language proficiency is unlikely (by itself) to help students meet the new challenges of writing in the academy; nor would such a programme be especially motivating as students would perceive it to involve 'more of the same' (p.10). The XYZ University 2010 needs analysis of students' views also found that students did not want to continue with a secondary school-like curriculum focusing on form but wanted a new challenge.

In addition to the Evans and Green (2007) study, two other large-scale investigations of students' English language needs at tertiary institutions in Hong Kong have been conducted (Hyland 1997; Littlewood and Liu 1996). These studies suggest that productive skills and acquisition of subject-specific vocabulary are the main language difficulties of the student participants, yet this body of research covered the four skills and was thus not specific to academic writing. Braine and McNaught (2007), in rationalizing the need for a Writing Across the Curriculum programme in Hong Kong universities, specifically researched writing provision at secondary and tertiary levels and found that the existing approaches to writing instruction were not effective in the Hong Kong context.

Evans and Morrison (2011) conducted a longitudinal study focusing on the challenges faced by first year students at Hong Kong universities. Their qualitative findings reveal four areas of difficulty in adjusting to the demands of university study: understanding technical vocabulary, listening to lectures, writing in an appropriate style and conforming to the conventions of the new academic discourse community. They also concluded that students from CMI schools struggle with the transition to university much more than their peers from EMI schools. At the ELC at XYZ University, the vast majority of students are products of CMI schools and have

graduated with a level 3 in the English component of the HKDSE school leavers' exam (the minimum passing level). The challenges mentioned in these studies (Hyland 1997; Littlewood and Liu 1996; Evans and Green 2007; Evans and Morrison 2011) are commonplace among these students.

The new Hong Kong Senior Secondary English Language Curriculum does not cover academic literacy skills, although the Curriculum and Assessment Guide (Secondary 4-6) (2007) lists various aims in the Writing, Reading and Language Development Strategies sections which seem to refer to the kinds of processes involved in writing from sources, for example "present and elaborate main ideas and supporting details through exemplifications, paraphrases, explanations", "evaluate critically views and attitudes"(p.25). It also includes reference skills, such as using "the library and the Internet regularly to collect information and develop research skills‖; information skills, such as identifying ―relationships … between the ideas expressed within texts"; and adapting ―materials, text-types, systems, etc., for supporting and illustrating various topics‖ (p.26). However, secondary teachers attending workshops offered by the ELC at XYZ University had continuously stressed that the more demanding of these processes can only be attempted with the highly competent students, and that in the majority of classrooms, where texts are used to inform writing, the genres tend to be the less formal ones such as book or film reviews or letters to the editor. Students have certainly not been instructed in the academic conventions of citation, and in some instances seem to have been encouraged in poor academic attribution practices. For example, in the HKDSE Listening and Integrated Skills paper students are rewarded for including key content points from a data file in their report but there is no requirement to attribute the source. Given that this exam is a university entrance test, this seems to indicate poor construct validity.

Questions on the HKDSE writing exam may occasionally elicit an analytic/persuasive genre similar to those required at university, but the content would be more personal than academic, for example: *Your school principal is thinking of reducing the number of PE lessons per week. Write a letter to the principal expressing your opinion towards this proposal*. It therefore comes as no surprise that many entering students rely on broad generalizations, sweeping statements and personal details as supporting details rather than rational, grounded ideas. New students also lack

experience in elaborating their arguments with logically ordered propositions. Many employ horizontal development, restating previous ideas rather than extending and developing an argument. There is also a tendency to adhere to the formulaic five-paragraph essay structure, with the inclusion of certain elements such as a counter argument merely because the model specifies it, rather than to introduce a relevant argument. In rater training sessions for EAP assessments, when discussing writer's stance and voice, teachers at XYZ University comment that while some students overstate their opinions, others struggle to develop a position. The need to moderate lines of argument and establish a scholarly and authoritative voice becomes even more challenging for many when they begin to incorporate source texts into their compositions.

New undergraduates also tend to 'play safe' rather than experiment with the written word. These students have been educated in an exam-oriented culture, in which secondary school teachers tend to adopt conventional approaches to writing instruction, assigning a topic and requiring students to produce a single draft in exam-like conditions. In many secondary schools accuracy prevails, teachers correct every grammatical error and the emphasis is on the score rather than on ways of improving the essay (Lee, 2012). This behaviour is evident in the frequency of memorized chunks of text and the reliance on formulaic language and model essays in the students' writing. The trend of memorizing model essays, although still evident, was particularly strong under the old A-level exam which had fewer question choices than the current HKDSE exam. Carless (2011) remarks that —the obsession with examinationsǁ (p.71) manifests itself in the thriving tutorial centre culture in Hong Kong where the vast majority of students attend classes after school in their final two years of secondary education. Tutorial centres claim they can provide exam tips and insights and much of this instruction focuses on model essays. In 1996 a particular tutorial centre made the headlines when tutors claimed they had inside knowledge of the A-level paper and misinformed 584 students, who subsequently received low scores for producing a memorized model essay (Bray, 2003).

Author 2 (2007) argues that the dominant exam culture in Hong Kong stifles a true learning culture as the test format becomes a model for curriculum, driving instruction, narrowing learning and distancing teaching from the real needs of

students. In her work with local secondary schools, Author 2 discovered that many teachers were afraid of implementing alternative methodologies, such as process writing, as they were worried about censure from department heads if students underperformed in exams. Another possible reason for teachers' heavy focus on accuracy and the writing product may be the influence of mother tongue instruction. Author 2 (2007) and Carless (2011) both point out that many Hong Kong teachers do not have basic teaching qualifications let alone training in ELT. In the absence of more informed or alternative approaches, teachers resort to the strategies used when they were students, thus perpetuating traditional approaches. Carless (2011) remarks that in traditional Chinese instruction, memorization and learning of the forms of the characters is prevalent and that this focus on form and perfection of the written product generally takes precedence over using the language as a real means of communication.

This traditional emphasis on the linguistic and structural features of texts and the reliance on model essays tends not only to inhibit students from experimenting with language, but also to prevent them from fully engaging with the topic by developing their own ideas and establishing their identity as an author. Finding the balance between their own voice and ideas from sources is challenging for first year university students who may see the two as conflicting demands (Groom, 2000). Naturally, these problems are not universal among our students, given the range of schools, teachers and life experiences from which they come; we do not wish to stereotype the students but to indicate the kinds of difficulties often encountered by students moving from secondary to university education in this context.

A new framework for EAP

The core English programme at XYZ University which has resulted from this period of curriculum and assessment development is one-year general EAP course, focusing wholly on reading and writing for academic purposes and introducing students to the academic literacy skills they need to cope in their subject courses. (This is an in-sessional programme, thus students are concurrently studying their major content courses.) The text-based curriculum is informed by genre-based pedagogy, emphasising guided analysis of authentic genre exemplars, and the

argumentative essay is the main writing genre and assessed task-type. It was considered important to address the construct validity concerns of the use of inauthentic independent writing assessments in university settings by introducing integrated (reading-writing) assessment. Figure 3 shows the assessments in the EAP programme with the analytic essay tasks highlighted. The course also includes a summary writing assessment and two reading assessments, but this paper focuses only on the analytic reading-into-writing assessment tasks, given their significance in the curriculum and their importance in students' academic work within the mainstream of the university.

Insert Figure 3 here

Reading-into-writing tasks

In EAP programmes generally it is increasingly recognised that much of the work that students do involves reading large amounts of text, both assigned and resulting from independent searches, and drawing on those texts to provide sound evidence to support, strengthen and further their own arguments. The decision to implement reading-into-writing assessments in the new EAP course at XYZ University to replace independent writing tasks represents the attempt to more closely mirror real-life language use in university disciplines. In academic study, reading and writing are inseparable, as students display their emerging knowledge of their subject area with reference to experts in the field. Thus writing-only tasks are increasingly seen as inauthentic (Gebril and Plakans, 2014). This process of switching between reading and writing to create meaning involves higher order sub-processes such as selecting, synthesizing, connecting and elaborating and poses immense challenges to students (Grabe and Zhang, 2013; Hirvela and Du, 2013).

Knoch and Sitajalabhorn (2013) define the construct of integrated writing as:

> tasks in which test takers are presented with one or more language-rich source texts and are required to produce written compositions that require (1) mining the source texts for ideas, (2) selecting ideas, (3) synthesising ideas from one or more source texts, (4) transforming the language used in the input, (5) organizing ideas and (6) using stylistic conventions such as connecting ideas and acknowledging sources (p. 307)

They add that rating scales used to grade such tasks need to fully capture all of the aspects of the construct being assessed by the integrated tasks, implying that measurement instruments for independent writing tasks are not adequate. The Assessment Team at XYZ University designed a term paper assessment, which is a take-home assignment requiring students to read around the topic and cite from their own and given sources. This kind of assignment is a more naturalistic, complex and multi-faceted task, similar to the writing of an actual dissertation (McCulloch: 2013). Plakans (2012) refers to this as the 'fidelity claim' of an integrated task.

In addition to the term paper assessment, two timed compositions were included in the EAP assessment framework, one at the end of the first semester, intending to provide both a mid-course performance indicator and formative feedback, and one at the end of the second semester. The EAP course, although providing a foundation in academic literacy skills, also serves a gate-keeping purpose and thus formal timed exams rather than purely coursework are deemed necessary. The inclusion of both timed reading-into writing assessments and take-home assessments provides opportunities for students to perform to the best of their ability under different conditions. Cumming (1997) remarks that the use of multiple writing tasks and the integration of reading and/or listening materials for assessing academic writing allows raters to make more informed decisions about student ability.

During the three-year development period, the ELC's curriculum and assessment teams worked closely to develop reading and writing materials and to produce writing assessments that would meet the requirements of the new focus on academic, integrated reading and writing tasks. In addition to the development of new task types and assessment papers, a key concern of the assessment team was with a new scoring instrument assessing the reading-into-writing construct. The development work and trialing took place from 2008 to 2011, with paid students invited to pilot the new tasks. The samples of work generated from these trials informed the development of the rubric. In the 2010 to 2011 academic year, a small-scale, year-long pilot of the complete course and assessment framework was done with five EAP classes. The course was fully operational the following year with a complete rollout of the programme and assessment framework for the full cohort of

students entering university in 2011 (although these were not HKDSE students as the new exam was introduced in the 2011-2012 year).

Development of the writing assessment instrument

It was clear from the outset that the new writing assessment instrument needed to satisfy both internal (EAP programme) needs, and external (university senior management) expectations. First, it needed to reflect the new focus in the curriculum on academic literacy rather than language proficiency. The argument that assessment in writing in academic contexts should probe not only the students' general linguistic proficiency but also their ability to handle academic content is a long-standing one (Hamp-Lyons 1991; Weir 1983). This is essentially the argument made by Hamp-Lyons and Kroll (1997), that assessments in academic contexts must reflect students' ability to use writing for knowledge-making as well as for knowledge-telling. Second, teachers in the ELC wanted an instrument that would form an integral part of the course and would be sufficiently detailed to provide valid diagnostic information for themselves and their students. But thirdly, there were also institutional expectations that the new assessment would be sufficiently reliable for upward-reporting to inform planning and reporting beyond the university. Part of this last expectation was, returning to the earlier discussion of the demand for explicit and comparable cross-university standards of English, that the assessment would produce outcomes, i.e., scores, that were aligned with a common metric. In the absence of any Hong Kong based common metric, and with the removal of the IELTS/CEPAS, the choice that management turned to was the Common European Framework of Reference.

The 2008-2011 development of the writing scoring instrument for the new EAP programme was the first time the ELC Assessment Team had produced a writing rubric that was empirically derived and that was aligned to an external standard. In the past, ELC scoring instruments had been developed based on teachers' expertise and experience and other good practice, not on actual performance. North and Schneider (1998) comment that the kind of practice previously employed in the ELC is commonplace: however, this time the performance criteria were grounded in close study of actual samples of student writing. Writing

samples on the newly developed writing test were obtained from a range of students at different proficiency levels. These compositions were read, discussed, re-read and discussed again, in a detailed iterative process by a group of experienced EAP teachers among whom were teachers with backgrounds in assessment and in discourse analysis. The aim was to progressively elicit and capture criteria that accurately reflect the features found in real student writing and the levels that are representative of the range of abilities in the programme (Darling-Hammond et al. 2013; Leki, Cumming & Silva, 2008). From this process a list of desirable qualities, which later became the key performance indicators and assessment criteria, started to take shape. As this set of target features found in actual student writing was developed, it was then used as a reference or guide when re-analyzing scripts to pinpoint differences between levels of performance. Gradually it became possible to put language to these levels, and the descriptors slowly emerged. Although the set of scripts had initially been chosen for analysis based on the A-Level grades of the pilot students, the actual features in the scripts themselves became more salient and starting grades were left behind. The process of writing descriptors for each domain and at each level was again a lengthy, iterative one: going backwards and forwards, discussing scripts and features they displayed and then reworking, combining and renaming the descriptors and assessment criteria until a group consensus was reached.

By the summer of 2011, the initial version of the scoring instrument was in reasonable shape, and feedback was sought from other teachers in the department in order to refine it further. Teachers were given sample scripts and instructed to grade them and give comments on comprehensibility and user-friendliness of the rubric. It was most important to have the acceptance of the wider ELC community as the rubric was to be used for large-scale assessments involving up to 2000 students and 40 teachers. The feedback and scores from teachers were discussed and the rating scale was adjusted accordingly. At this time, no feedback was sought on the alignment with IELTS / CEFR or the EAP passing standards, and the focus was solely on the feasibility and effectiveness of the rubric as an assessment tool for the particular construct. The scoring instrument that emerged from this process comprises six domains and six score

levels, as shown in Figure 4.

As previously mentioned, an internal benchmarking exercise of EAP writing standards was carried out to facilitate teachers of the new EAP programme in contextualizing and interpreting the levels on the new rubric. The outcome of this alignment is shown in Table 2. It can be seen that three EAP grades (B, C and D) are compressed into one IELTS or CEFR level, and that the difference between the internal grades of D and A is only 6 points on a 36 point rubric. Such compacting of levels would require an extremely finely-grained rubric to distinguish between performances. This is the consequence of the attempt to create actual equivalences between levels on a full spectrum international set of standards such as the CEFR, or even the internationally-recognised IELTS, while still delivering an internal assessment that captures, reports, and allows teachers to feed back relatively small changes in writing proficiency within the range of the population being taught.

Trialling and validation of the new instrument

The official rollout of the new programme across the whole department took place during the 2011-2012 academic year. This was both a live implementation and a validation study of the assessments and the 6X6 scoring instrument, but with HKALE graduates as the HKDSE was only implemented in 2012.

Assessment 1

The first live data on student performance became available in December 2011 with Assessment 1, the mid-semester test. This assessment was a 600-word academic essay, with assigned sources, to be written in 1 hour 45 minutes. The design intentions were to familiarize students with the format of the final exam and to provide a feedback-rich assessment mid-way through the course so that poorly performing students would have a constructive idea of the areas they needed to focus on in the second semester. Carless (2007) refers to this as a feed-forward

assignment as there is a built-in opportunity for students to put the feedback to immediate use. See Figure 3 for assessment framework.

The rating of this essay by 41 EAP teachers was the initial large-scale application of the 6x6 rubric. Owing to the large cohort of students (1547), scripts were single-marked. Spot checking was done to identify erratic raters for future training purposes but actual scores were not changed. Raters had attended familiarisation sessions on the new rubric prior to the semester and also a rater training session before the mid-course test using live scripts. A major part of the Assessment Team's work had been the creation of a booklet of exemplar scripts at each level of performance with detailed rationale for scores and alignment to CEFR, IELTS and the soon-to-be defunct HKALE (as teachers and students were familiar with the A-level standards). The team hoped that the provision of exemplar scripts at the start of the semester would help teachers understand the criteria and levels in the new rubric and ultimately would make the rubric more transparent to students.

The results of this initial large-scale use of the 6x6 instrument were, bluntly, shocking and disappointing. There was a 60.4% failure rate based on the pass mark of 20/36 (as shown in Table 2), which had been set with reference to IELTS and CEFR, and the top two levels (5 and 6) on the rubric were rarely used by raters.

Analysing these results according to both score distribution and grade distribution provides a very interesting picture of student performance (see Figure 5).

Insert Figure 5 here

The distribution of students' raw scores out of 36 shows an almost bell curve. However, once the scores have been converted to grades based on the cut scores set during the alignment exercise, the bar chart looks very different, as the number of students increases dramatically from grades A to F, i.e., is very strongly skewed towards weak scores. If viewed as an independent population,

15

after one semester of EAP this group of Hong Kong students is almost normally distributed, yet the grade distribution gives a completely different image of the same population.

Referring to the EAP alignment in Table 2 helps to explain the distribution more clearly. The December 2011 data collection showed that far too many students were not achieving the passing standard of 20 out of 36, in fact only 19% of the total population scored higher than a D. It appears that the rubric was not delicate enough to distinguish the many levels within the CEFR B2 band, as EAP grades A to F are compressed between 19 -- 26 on a 36 point scale, meaning the difference between each grade is statistically negligible. A difference of just one level in two of the six domains on the writing rubric (Figure 4) would translate into a whole grade difference.

Informal conversations with raters unveiled other major concerns with the scoring instrument which may have contributed to the high rate of failures. Many reported that the level 3 descriptors read like a pass, yet a score of 3 across all six domains did not reach the EAP passing standard of 20. This was possibly due to the misleadingly positive nature of the 'Can Do' statements. It was obvious that some kind of urgent intervention was necessary before embarking on the 2nd semester. Fortunately the writing assessment in the December test had only contributed 10% towards the total EAP grade, but looking ahead to the term paper and final exam, which carried more weight, it was politically unacceptable to have such high failure rate. There were admittedly some reliability concerns based on certain raters' scoring patterns, heightened by the fact that the essays were not double-rated, but the team were convinced that the passing requirement had been set too high. Although lowering the standard would have been a quick (and dirty!!) fix, this was not a viable option half-way through the year as the expected standards / criterial levels of performance had been made publicly available.

After endless internal troubleshooting, advice was sought from an impartial writing assessment expert who had not previously been involved in the EAP curriculum, assessment or rubric development and benchmarking. The decision was made to 'open up' the rubric between levels 3 and 4 by adding a 3.5 band.

Level 3 on the rubric was the most widely used band for each criteria, yet with six domains and a passing standard of 20, this would result in a fail. It was hoped that the new 3.5 band would fit some of the students who were displaying elements of the performance required by both levels 3 and 4, and that this would thus push these students up to a grade D, the minimum passing grade. In effect the 6x6 became a 6x7. This meant there were even more cells on an already ‗busy' rubric but since moving forward with such a high percentage of failures was not an option, this seemed like the best solution.

Assessment 2

The second large-scale use of the 6x6 rubric was the rating of the term paper assessment in the second semester of EAP (Semester B, January - March 2012). The term paper was a high stakes test at 25% of the overall score. The task required students to write a 2000-word essay based on assigned sources and also on their own source research. Drafts were submitted initially for peer feedback and subsequently for teacher feedback. Each student attended a consultation before revising their work and submitting it for final grading. There were 2000 students in this cohort and 45 teacher-raters. Scripts were again single rated due to the lack of resources but this time, teachers did not rate their own students' papers and as with Assessment 1, spot checking was done.

As is the usual practice in the ELC, all raters attended a training session in which five scripts were analysed and discussed together. Clear rationale was provided for the scores for each domain and sample scripts at each performance level were distributed. Due to the heavy weighting of this assessment and the observation of some erratic scoring patterns after the mid-course test, there was a need to be more stringent this time. Thus, the Assessment Team stipulated that all raters had to take four scripts away to grade on their own. Raters were required to submit their scores to the team and were only given their term paper scripts to grade once their rating of the four training scripts was satisfactory. For the majority of raters, the initial set of scores submitted was acceptable, however 4 of the 45 teachers had to grade the scripts again and resubmit their scores.

The good news was that this assessment yielded a much lower failure rate (38%), but in all honesty it was impossible to discern if this was primarily because of the new 3.5 band. As this assessment followed a process writing approach with two feedback cycles and three versions being written, we would expect a better performance than in a timed-essay, and also students had received a few more months of EAP instruction. It was observed that the 3.5 band had been extensively used by raters and when interviewed they reported that the band was useful and made decisions at that level of performance easier. Similar to the previous assessment, levels 5 and 6 were rarely used and there were still concerns about rater reliability when looking at the score distribution among raters.

Final exam

The end of the 2011-2012 academic year brought the 3rd trial / validation of the 6x6 rating scale with the final writing exam (end of Semester B, April 2012). This task was identical to the writing task in Assessment 1 (a 600-word academic essay, with assigned sources, to be written in 1 hour 45 minutes) and it contributed 25% to the overall EAP grade. There were 1550 students and 25 raters, who had all been teaching EAP. No changes had been made to the rubric, but there was one more round of rater training. Scripts were double-rated for the first time with 3rd rating being done when there was a difference of 3.5 or more between the first two raters.

Although the positive downward trend in the failure rate continued, this was not significant and at 32%, it was still politically unacceptable. Despite continued attempts to train raters, inter-rater reliability concerns persisted, with almost 30% of scripts needing 3rd marking based on a difference of 3.5 between first and second raters. Serious cases of aberrant marking were noticed and scripts from three raters had to be entirely re-graded.

User feedback in the form of questionnaires and teacher interviews was collected and it was clear there was a constant tug between the formative and summative demands of the rubric. Teachers overwhelmingly favoured the 6x6 rubric as a feedback tool as it provided explicit and detailed feedback. However, the vast

majority reported that it was too demanding for formal scoring due to the level of detail, the large number of cells, the idiosyncratic separation of Grammatical Accuracy and Syntactic Complexity and the bleeding over of certain features between domains.

Lessons learnt

A year of rigorous trialling had led the assessment team to the realization that they were not confident in the decisions being made based on this rubric. They felt it was failing to give a valid and satisfactory representation of the students' EAP writing performance. Therefore it was decided to undertake a significant revision in parallel with revisions to the EAP curriculum. There was a need to streamline the assessment tasks in order to reduce the marking load for teachers and thus allow for more double-rating. Rather than having an identical assessment at the end of the first semester and the final exam, it was agreed that the assessment should more closely follow the curriculum, with each assessment building on the previous one, either by requiring newly-taught skills, providing more source texts or requiring a higher word count.

It was clearly essential to revisit the original design intentions looking at the criteria, the performance standards and the alignment issue before the start of the 2012-2013 academic year. After much disagreement and a reluctance from some team members who had developed a strong bond with the 6x6 and thought it should be granted more time to prove itself given the amount of work which had been put into the development, three key decisions were made:

i. To work with the 6x6 in the curriculum as a feedback instrument because of its value for providing detailed feedback and diagnosis.
ii. To transform the 6x6 into a simpler, more teacher-friendly instrument for assessment purposes → the 4x6 (see Figure 6).
iii. To leave reference to CEFR/IELTS/HKDSE behind.

Place Figure 6 here

Putting the lessons into practice

Development of the 4x6

To address the issue of the 6x6 rubric being too complicated for summative assessment purposes, the instrument was reorganized into a simpler 4-trait version comprising Task Fulfillment, Discourse Competencies, Language Competencies and a newly-developed trait, Source Integration. The resulting 4x6 rubric encompasses the same key aspects of academic literacy as the 6x6, while focusing on specific areas where Hong Kong university entrants are typically weak.

This newly-adapted rubric went through similar stages of trialling to the original 6x6. In course evaluation meetings, teachers have reported feeling more comfortable with a simpler rating scale and there is less tension at rater-training sessions. As we now use multi-faceted Rasch measurement (MFRM) analyses to understand patterns of rater severity or leniency, and rater effects generally, there is much less concern to have raters agree, and consequently more harmony between raters. (Authors in press). A MFRM analysis conducted on the data from the final exam in April 2014 showed that the mean scores for raters for all domains are within the range of 2.83 to 3.29 and quality control fit statistics were all within the recommended parameters. The Infit mean square values, which give information about how consistently raters used the scale, ranged from .74 to 1.38: values between. 0.5 and 1.5 are seen as ideal.

Delinking from external standards

While the 4x6 addresses the same key competencies of academic writing as the 6x6, the levels and passing standard were set based on sample scripts from a trial without needing to worry about alignment with external metrics. The XYZ University assessment descriptor band names (Fail, Marginal, Adequate, Good, Excellent) were adopted for each level to contextualize the EAP-internal standards within the assessment framework of the wider university. Setting our own standards based on the local students' actual performance gave the assessment team confidence in judging how well the students are grasping key

attributes of academic literacy and also in their pass/fail decisions. The team also gained confidence from work by Fulcher (2004) who in the early days of the CEFR expressed caution about mediating between test scores and linking tests which have been designed for different purposes. More recently the CEFR has been widely adopted by providers in New Zealand to benchmark a new suite of English language qualifications (NZCEL), although Read (2014) concluded there are few indications that the CEFR will play a direct role in language provision or make any practical difference to language assessment in New Zealand. Similar caution is being expressed elsewhere. In a series of studies to assess the appropriateness of using standardized test scores for making ESL placement decisions within the context of a US university, Kokhan (2014) suggests that using external test scores is not a reliable method of judging language understanding and ability and often results in a significant number of misplaced students.

Discussion

The effects of delinking the instrument
Throughout the different stages of development, trialling, revising and familiarizing the whole group of EAP teachers with the 6x6 rating instrument, there had been discussions within the assessment team about **why** it was necessary to align the local instrument to an external standard, especially one as distant from the context of Hong Kong tertiary education as the CEFR seems to be. The external consultant, with the benefit of distance from the problem, was able to see that the need for a stable metric for comparison across Hong Kong, and potentially beyond, was not a need established by the ELC but one brought in from university senior management. Once a decision had been made to delink the metric part of the scoring instrument from the CEFR, the assessment team were able to look at the writing produced by these students—students who only one year before, in the old structure, would have been entering into a final school year rather than a university course, for its own characteristics and for the patterns of strengths and weaknesses displayed by local students who have met the university entrance requirement, even if only minimally.

This meant that the EAP programme regained control of its passing standard and

was not driven to describe large numbers of students as 'failing' because of the alignment with an inappropriate choice of benchmark. With the revised curriculum and streamlined assessments, it meant the teachers in their classrooms, were able to focus on the key activities of teaching and supporting students' learning through positive and staged feedback. The 6x6 scoring instrument was retained for use in providing detailed feedback to students, for teachers who preferred to work from the more detailed document to provide diagnostic feedback, rather than giving personal feedback.

However, this of course meant that it was not possible to claim alignment nor to assert a CEFR-benchmarked 'average' for XYZ university's EAP programme leavers. However, with the support of the Centre Director and the increased awareness of senior management that alignment to the CEFR inevitably meant that formal score reporting would not look impressive, this change was allowed to stand, at least for the present.

The ethics of external benchmarking

This experience has led us to ask ourselves the wider question of what the ethics of external benchmarking may be when applied within a programme. EAP teachers accept that students have been and will probably be assessed on largescale tests that comprise, or are aligned with, an internationally-recognised standard. In Hong Kong EAP teachers have long been aware of the role of the IELTS in their students' future lives and careers. But when the external benchmark exerts such influence on how students are assessed internally, it becomes important to ask whether this influence is a good thing. Should our own students be judged on IELTS standards even when not sitting the IELTS? Should they be judged against an even more remote set of standards such as the CEFR? We have seen in this paper that working to external standards showed our students to be failures, despite the fact that they had never been asked to meet the standards at which they failed. But neither the CEFR nor the IELTS are really the problem here: the problems are more a consequence of government officials' and university managers' lack of proper understanding of assessment tools such as the CEFR or IELTS. In particular these bureaucrats don't seem to be aware of how wide the CEFR bands are relative to any local instruments which are targeted to specific populations: this is well illustrated in

Table 2, where we see that most score levels in the internal benchmarking aligned to the same CEFR level (B2) and the same IELTS level. We believe that if Hong Kong needs a 'common standard' for local political reasons, it should create a local one. Hong Kong language centres are assessing an academic literacy construct that is based on current understanding of how English is actually used in academic contexts, and that is quite well-aligned with research in the field. Although it is claimed that IELTS is suitable for people planning to study in English, the current IELTS Academic is only loosely aligned with academic language, and the CEFR not at all.. If Hong Kong students want to compete for study places in other countries, large-scale tests such as IELTS and the TOEFL can judge them against a wider international student body. But these purposes are not the same.

We have also seen the problems caused for curriculum development and materials preparation, as teachers confront the reality that the 144-hour EAP course cannot bring the weaker students—students whom the universities have accepted—from their entry level to the required standard no matter how good the materials or the teacher. We have had to ask ourselves which is more important: student progress or external accounting? For teachers there can only be one answer. For programme leaders it is more difficult. The next question therefore becomes: who decides? Should it be the team -- practising EAP teachers with happenstance responsibility for assessment within the curriculum and the institutional requirements? Or should it be the management, who among other things are responsible for the status of the institution in the public perception? At this point we can ask the question rhetorically, since there is now a writing assessment that is working well and that has stood the test of two years of full implementation.

Conclusions

There were several reasons why this process was so difficult. First, too much was happening all at once. The revised curriculum, the new assessment tasks and the rubrics were all being developed at the same time as the standard setting / benchmarking exercise took place. It would have been better to have a stable instrument which had been validated before embarking on the alignment

exercise. Second, the demands and expectations from senior management for the 'success' of these younger and less experienced students were too great, indeed unrealistic; they were also set without any understanding of what each test can (and cannot) do. Thirdly, there was not nearly enough knowledge and understanding of the CEFR among the staff involved in the project: the CEFR manual describes detailed and systematic procedures which should be followed during an assessment development process. This document is extremely useful, but it was not published until this project had already started, and the implications of the attempt at linking were not grasped until very late. Furthermore, any multiple trait scoring instrument requires multiple standard setting, i.e., the alignment of each trait scale independently to the target. Such complexity in turn requires that all assessors (and in an EAP context, all teachers too) fully understand and are aligned with the instrument and its associated external yardsticks.

A way forward?

We see two possible ways forward, but in rather opposite directions. First, XYZ University—and perhaps the Hong Kong universities as a joint body--could consider **a dynamic criteria mapping project** (Broad, 2003) to achieve locally-based validity and reliability. This would engage teachers (and, ideally, students) in tying curriculum and assessment together with the aim of developing a more fine-grained understanding of what specific writing skills each assignment calls for, and establishing locally-relevant criteria to assess those demands (Authors, in press). Second, if an alignment to the CEFR is demanded, it should first of all be **carried out directly from the EAP programme assessments to the CEFR** and not indirectly via the HKDSE or the IELTS. But perhaps a middle of the road solution would be possible: Hong Kong applied linguists and EAP specialists could build a shared set of assessment tools on a common local yardstick: this is rather like what projects such as the DELTA are working towards but at a more general level. As these shared local tools emerge, they could be fitted together in order to gradually **create a Hong Kong Framework of Reference for English** *(HKFRE).* To succeed in this, the development team(s) would need much more familiarity with CEFR (or similar) principles, tools, and procedures for

relating individual tests and tasks to the common structure than any single institution's EAP team would normally have. The CEFR is, as its name states, a *framework* and not a test. Its strength is in its advisory and descriptive richness, and teams should draw on these. Implementation would need to be gradual and all aspects should be validated and stable before formal alignment is attempted.

References

Authors (In press).

Author 2 1991.

Author 2 2007.

Author 2 and A.N. Other (1997).

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Press.

Bray, M. (2003). *Adverse effects of private supplementary tutoring: Dimensions, implications and government responses.* (Ethics and Corruption in Education Series). Paris: IIEP- UNESCO.

Broad, B. (2003) *What we really value: Beyond rubrics in teaching and assessing writing.* Logan: Utah State University Press.

Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International* 44,1: 57–66.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings.* New York: Routledge.

Cumming, A. (1997). The testing of writing in a second language. In D. Corson & C. Clapham (Eds.), *Language testing and assessment: Volume 7 Encyclopedia of language in education* (pp.51-63). Dordrecht, The Netherlands: Kluwer.

Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for highquality assessment.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Field Code Changed

38

Evans, S. & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes 6*: 3–17.

Fulcher, G. (2004). Deluded by artifices: The Common European Framework and harmonization. *Language Assessment Quarterly, 1,* 4: 253-266.

Gebril, A. & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing, 21*: 56-73.

Grabe, W., & Zhang, C. (2013) Reading and Writing Together: A Critical Component of English for Academic Purposes Teaching and Learning. *TESOL Journal, 4,* 1, 9-24.

Groom, N. (2000). A workable balance: Self and sources in argumentative writing. In Mitchell, S. and Andrews, R. (Eds.). *Learning to Argue in Higher Education.* Portsmouth, NH. Boynton/Cook.

Hirvela, A., & Du, Q. (2013). ―Why am I paraphrasing?‖: Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes, 12,* 2: 87-98.

Hong Kong Examinations and Assessment Authority (2007). *English Language Curriculum and Assessment Guide.* Hong Kong: Curriculum Development Council.

Hyland, K. *(2002). Teaching and researching writing.* Harlow: Pearson Education.

Knoch, U. & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focussed definition for assessment purposes, *Assessing Writing* 18, 300–308.

Kokhan, K. (2014). Examination of the appropriateness of using standardized test scores for English as a second language (ESL) placement. Unpublished doctoral dissertation, Graduate College of the University of Illinois at Urbana-Champaign.

Lee, I. (2012). Genre-based teaching and assessment in secondary English classrooms. *English Teaching: Practice and Critique,* 11, 4, 20-136. http://education.waikato.ac.nz/research/files/etpc/files/2012v11n4art8.pdf pp. 120-136

Leki, I., Cumming, A. & Silva, T. (2008). *A Synthesis of Research on Second*

Language Writing. London: Routledge.

Lui, E. (2009). The standard adopted by university language teachers: Which one? A case study. Seminar at the Language Testing Research Forum, CRELLA, University of Bedfordshire, June 2009.

Lui, E., Wong, R., Huckstep, N. & Yuen, W. (2010). English Language Centre Needs Analysis Report. Hong Kong: Author.

Lockwood, J. (2013). The Diagnostic English Language Tracking Assessment (DELTA) writing project: a case for post-entry assessment policies and practices in Hong Kong universities. Papers in Language Testing and Assessment, Vol 2.

McCulloch, S. (2013). Investigating the reading-to-write processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study. *Journal of English for Academic Purposes*, 12, 2:136 – 147.

North, B. & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 2::217-262.

Plakans, L. (2012). Writing integrated items. In G. Fulcher and F. Davidson . (Eds.), *The Routledge Handbook of Language Testing*. Routledge

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing,* 17(1), 18–34.

Read, J. (2014). Influence from afar: The CEFR and a New Zealand tertiary-level qualification. Presentation given at the EALTA Conference June 2014. Retrieved 1-8-14 from:

http://www.ealta.eu.org/conference/2014/presentations/John%20Read%20 EALTA%202014.pdf

Weir, C. (1983). *Identifying the language problems of overseas students in tertiary education in the United Kingdom*. Unpublished PhD thesis, University of London.
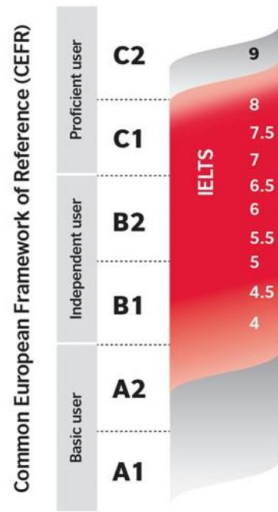
Figure 1: CEFR-IELTS concordance

Figure 2: Timeline of key events in EAP Curriculum and Assessment Development

See separate file

Table 1: Concordances between HKALE/IELTS, HKDSE/IELTS and IELTS/CEFR

HKALE Use of English (AS Level)[1]	IELTS	HKDSE English Language[2]	IELTS	CEFR[3]
A	7.41 - 8.30	5**	7.51 – 7.77	C1
B	6.92 - 7.40	5*	7.16 – 7.32	C1
C	6.51 - 6.91	5	6.81 – 6.99	B2
D	6.03 - 6.50	4	6.31 – 6.51	B2
E	5.40 - 6.02	3	5.48 – 5.68	B2
		2	4.79 – 5.07	B1

[1]http://www.hkeaa.edu.hk/en/recognition/benchmarking/ce_al/ielts/

[2]http://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/

[3]http://www.ielts.org/researchers/common_european_framework.aspx

Table 2: Outcome of internal benchmarking exercise

Rough Equivalences between standardized test scores and EAP				
HKALE	IELTS	CEFR	EAP Grade	EAP Score Max: 36
C	6.5	B2+	A	26+
D	6	B2	B	24-25
			C	22-23
			D	20-21
E/below	5.5/below	B1+/below	F	<20

Assessment	%	Details	Dates
Assessment 1: Mid-course test (20%)	10	Reading, Usage and Summary	End of Semester A
	10	600-word timed analytic essay using assigned sources	
Assessment 2: Term Paper (30%)	5	Summary of assigned text	Start research and make a plan in Semester A and complete in Semester B
	25	2000-word research paper on a given theme using assigned and own sources (process writing approach)	
Final Exam: (50%)	25	Reading, Usage and Summary	End of Semester B
	25	600-word timed analytic essay using assigned sources	

Figure 3: EAP Assessment Framework (piloted in 2011 to 2012) with tasks relevant to this study highlighted

FIGURE 4: A section of the 6X6 writing rubric

I E L T S	C E F R	Mark	TASK FULFILLMENT & CONTENT	ACADEMIC REGISTER & STANCE	ORGANISATION	SYNTACTIC COMPLEXITY	GRAMMATICAL ACCURACY (EXCLUDING SYNTAX)	TASK-RELATED VOCABULARY
			1. Responding to the given prompt: answering the question relevantly 2. Content: grounded arguments; evaluation of source texts; depth of detail; persuasive support	1. Academic register/audience awareness: acknowledgment of sources; referencing style; formal language (absence of spoken language); detached tone; appropriate use of hedging for academic writing 2. Stance: opinion/attitude of the writer on the subject matter and on cited sources	1. Structure and coherence: writing structure; transitions and flow within/ between paragraphs and between introduction and body; paragraph unity; topic sentences as appropriate 2. Cohesion: lexical cohesion and grammatical cohesion	Accuracy and variety in syntactic structures and sentence types; absence of syntactic errors like fragments or verbless sentences; use of modifiers to describe/clarify the meaning of the subject or the object	Accuracy level (syntax errors not included); impact of errors on readability	Accuracy (vocabulary errors-- collocation, word form and spelling), precision (absence of word choice errors) and variety (AWL, UWL/ technical words relevant to the task) in vocabulary use
6.5	B2+	6	• Can respond to the prompt given and fulfill the task with exemplary performance. • Can provide content sufficient to the task all the time.	• Can use an academic register to appropriately convey the writer's ideas to the reader all the time. • Writer's stance is clear to the readers most of the time.	• Can structure the writing clearly and logically in general. • Can use a variety of lexical and grammatical cohesive devices effectively almost all the time.	Can construct a variety of syntactic structures and sentence types correctly and use modifiers effectively all the time.	Can write with a high degree of accuracy though with negligible errors.	Can use a variety of task-related vocabulary accurately and precisely almost all the time.
6	B2	5	• Can respond to the prompt given and fulfill the task completely. • Can provide content sufficient to the task almost all the time.	• Can use an academic register to appropriately convey the writer's ideas to the reader almost all the time. • Writer's stance is clear to the readers some of the time.	• Can structure the writing clearly and logically in general. • Can use a variety of lexical and grammatical cohesive devices effectively with negligible problems in cohesion.	Can construct a variety of syntactic structures and sentence types correctly and use modifiers effectively most of the time.	Can write with a high degree of accuracy though with noticeable errors which have no impact on meaning and readability.	Can use a variety of task-related vocabulary accurately and precisely most of the time.
6	B2	4	• Can respond to the prompt given and fulfill the task well. • Can provide content sufficient to the task most of the time.	Can use an academic register to appropriately convey the writer's ideas to the reader most of the time.	• Can structure the writing clearly and logically in general with negligible problems in coherence. • Can use a variety of lexical and grammatical cohesive devices effectively though with a few problems in cohesion.	Can construct a variety of syntactic structures and sentence types correctly though with repetitive sentence patterns. Can use modifiers effectively most of the time.	Can write with a relatively high degree of accuracy though with noticeable errors which may have some impact on meaning and readability.	Can use a variety of task-related vocabulary accurately and precisely more often than not, maybe with a few instances of over-reliance on vocabulary from source texts (if applicable).
6	B2	3	• Can respond to the prompt given and fulfill the task adequately. • Can provide content sufficient to the task more often than not.	Can use an academic register to appropriately convey the writer's ideas to the reader more often than not.	• Can structure the writing clearly and logically in general though there may be a few minor problems in coherence. • Can use a variety of lexical and grammatical cohesive devices effectively though with a few problems in cohesion.	Can construct a variety of syntactic structures and sentence types correctly though with repetitive sentence patterns. Can use modifiers effectively some of the time.	Can write with a moderate degree of accuracy though with many noticeable errors which have some impact on meaning and readability.	Can use a variety of task-related vocabulary though not always accurately with a few instances of over-reliance on vocabulary from source texts (if applicable). Collocation errors are evident.

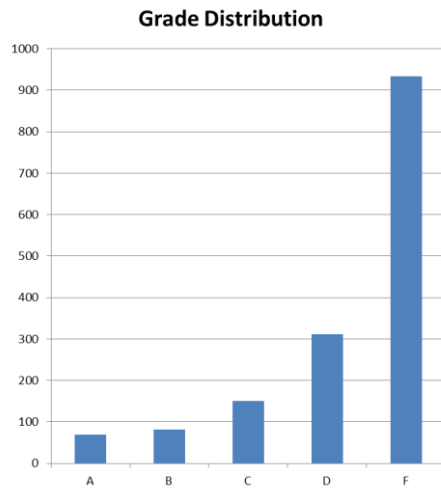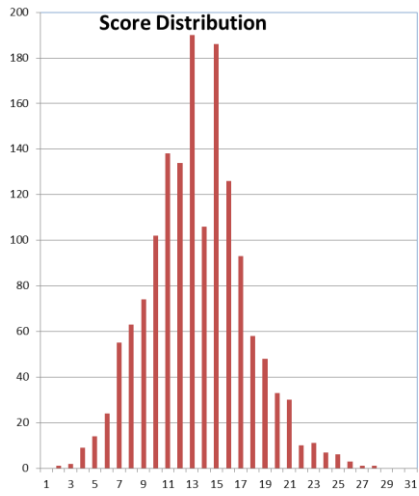# Figure 5: Assessment 1 overall results: distribution by score and grade

Figure 6 The 4x6 EAP Writing Assessment Descriptors
See separate file