

The beneficial washback of the school-based assessment component on the speaking performance of students

校本評核對學生說話能力表現的正面影響

LEE Wong Wai, Christina

Hong Kong Examinations and Assessment Authority

Abstract

This paper aims to show that the implementation of school-based assessment (SBA) has proved to have positive impact on the performance of students in the public oral examination. An SBA component was introduced to the Hong Kong Certificate of Education Examination English Language Examination in 2007. This consists of a reading / viewing programme where students need to read / view texts, write up some comments and personal reflections, and then take part in a discussion with classmates on the texts they have read / viewed, or make an individual presentation and respond to questions. The assessment is based on the student's oral performance. The 2007 experience has shown that the speaking ability of students can be reliably assessed in school by their own teachers. Statistically, the SBA component has proved to be as reliable as the speaking examination. The beneficial washback of SBA can be seen in the results of the speaking examination. Candidates from schools that submitted SBA marks had a lower absentee rate than candidates from schools not submitting SBA marks. They also performed better in the speaking examination.

Keywords

school-based assessment; Hong Kong Certificate of Education Examination (HKCEE) English Language; speaking examination; beneficial washback / backwash; speaking performance

摘要

本文旨在顯示推行校本評核證實對學生在公開考試的口試表現有正面影響。香港中學會考英國語文科在2007年引進校本評核，學生須閱讀或閱覽各種文本，包括小說類及非小說類的書籍及影片，並寫下評論及個人感想，然後與同學進行小組討論，或作個人短講並回答提問。旨在評核學生的說話能力。根據2007年的體驗，老師能可靠地評核學生的說話能力。統計資料亦顯示校本評核和口試的信度相若。校本評核的正面影響可見於公開口試的成績，於進行校本評核的學校就讀的考生缺席率較低，成績也較佳。

關鍵詞

校本評核，香港中學會考英國語文科，口試，正面影響，說話能力表現

Background

The Hong Kong Certificate of Education Examination (HKCEE) is taken by students in Hong Kong at the end of five years of secondary education. Examinations are offered in 39 subjects, mostly with equivalent English and Chinese versions, to around 100,000 candidates each year. The examinations assess candidates' achievement of the learning targets and objectives of the teaching syllabus promulgated by the Curriculum Development Council. The examinations are taken after a two-year course comprising Secondary 4 and Secondary 5 (S4 and S5).

A new HKCEE English Language syllabus including a school-based assessment (SBA) component was introduced in 2007 in order to align

assessment more closely with the English Language teaching syllabus published by the Curriculum Development Council in 1999 as well as the new Senior Secondary curriculum to be implemented in September 2009. The SBA component seeks to provide a more comprehensive appraisal of learners' achievement by assessing those learning objectives which cannot be easily assessed in public examinations while at the same time enhancing the capacity for student self-evaluation and life-long learning. The SBA, like the rest of the HKCEE English Language Examination, adopts a standards-referenced assessment system which seeks to recognise and report on the full range of educational achievement in Hong Kong schools. Table 1 outlines the examination syllabus.

Table 1: 2007 HKCEE English Language Examination

Component		Weighting	Duration
Public exam	Paper 1A - Reading	20%	1 hour
	Paper 1B - Writing	20%	1 hour 30 minutes
	Paper 2 - Listening & Integrated Skills	30%	2 hours
	Paper 3 - Speaking	15%	12 minutes
School-based assessment		15%	

The SBA component consists of a reading / viewing programme where students read / view three texts (“texts” encompass print, non-print, fiction and non-fiction material) over the course of two years, keep a log book of comments / personal reflections, and then take part in a discussion with classmates or make an individual presentation on the books / videos / films that they have read / viewed, and respond to questions from their teacher, which will be derived from the student’s written notes / personal responses / comments in their logbook. The assessment is based on the student’s speaking performance, that is, the reading / viewing / writing will only serve as the means to this end and the specific content of the texts (i.e. names and places, story lines, other factual information etc.) is not directly assessed as such.

Teachers are advised to develop the SBA component as an integrated part of the curriculum, not as a “separate” examination paper. Students should be encouraged to keep copies of the records of their

own assessments and regularly review their progress. Teachers should use the assessment activities not only to make judgments about student standards (a snapshot of students’ achievement to date), but also to give feedback to students about specific aspects of their oral language skills so that they can improve for the next assessment. The SBA component can be valuable preparation for students for their external HKCE examination, especially for the reading and speaking papers, as many of the skills required are the same.

The SBA component is worth 15% of the total English subject mark. In S4, teachers need to undertake at least one assessment of students’ group interaction or individual presentation skills and report one mark at the end of the school year. In S5, they need to again undertake at least one assessment of students’ group interaction or individual presentation skills, and report one mark at the end of S5. These requirements are summarised in Table 2.

Table 2: SBA Requirements

Requirements	S4	S5	Total
Number and type of texts to be read / viewed	One or two texts	One or two texts	Three texts, one each from three of the following four categories (print fiction, print non-fiction, non-print fiction, non-print non-fiction)
Number and timing of assessment tasks to be undertaken	One task, group interaction or individual presentation, to be undertaken during the second term of S4	One task, group interaction or individual presentation, to be undertaken anytime during S5	Two tasks, each on a text from a different category
Number, % and timing of marks to be reported	One mark reported at the end of S4	One mark reported at end of S5	Two marks, 15% of total English subject mark

An SBA handbook is published and distributed to schools to help teachers understand the rationale behind the introduction of SBA, and to provide guidelines regarding possible assessment tasks, assessment criteria and administrative arrangements.

SBA implementation issues

Original proposals regarding the introduction of the new language syllabus, including the details of the SBA component, were favourably received by schools and teachers when they were consulted in 2003 and 2004. However, as the schools started implementing the new syllabus with their S4 students in September

2005, a number of concerns were raised regarding the SBA, in particular concerns about workload, fairness, authentication of student work and teacher readiness. In April 2006, after a series of consultation seminars and a comprehensive survey of all schools, school councils and professional bodies, modifications were made to the design of the SBA component and the implementation schedule.

A three-year phase in period was introduced to accommodate variations between schools with respect to the optimum time to implement SBA. Schools can choose among three options. Details of the implementation schedule are shown in Table 3.

Table 3: Three-year Phase-in Implementation Schedule

Year	Options for schools
2007	1. Submit SBA marks for feedback and to contribute 15% of final subject result; or 2. Submit SBA marks for feedback only; exam results to contribute 100% of final subject result; or 3. Not submit SBA marks; exam results to contribute 100% of final subject result
2008	1. Submit SBA marks for feedback and to contribute 15% of final subject result; or 2. Submit SBA marks for feedback only; exam results to contribute 100% of final subject result
2009	1. Submit SBA marks for feedback and to contribute 15% of final subject result

SBA in the 2007 examination

Schools were asked to indicate their choice in October 2006 when they registered their students for

the 2007 public examination. Approximately one-third of the schools Option 1. Table 4 shows the number of schools and candidates involved.

Table 4: Number of schools and candidates choosing each option

Choice	No. of schools	Percentage (%)	No. of candidates	Percentage (%)
Option 1 (Yes)	199	34	31,875	43
Option 2 (Trial)	125	22	20,945	28
Option 3 (No)	254	44	21,388	29
Total	578	100	74,208	100

The 2007 examinations were conducted in May and June and schools choosing Options 1 and 2 submitted their SBA scores at the end of the two-year course, in April 2007.

Statistical moderation of SBA scores

One of the major concerns expressed by stakeholders, in particular parents and school teachers, is that SBA may not be a fair way of assessing student performance because teachers will conduct different teaching and assessment activities and schools will have different assessment plans to cater to the needs of their students. In order to ensure the fairness of SBA, the HKEAA uses statistical methods to moderate the SBA marks submitted by schools.

Teachers know their students very well and thus are best placed to judge their performance. In consultation with their colleagues, they can reliably judge the performance of all students within the school in a given subject. However, when making these judgments, they are not necessarily aware of the standards of performance across all other schools. Despite training in carrying out SBA and even though teachers are assessing students on similar tasks and using the same assessment criteria, teachers in one school may be harsher or more lenient in their judgments than teachers in other schools. They may also tend to use a narrower or wider range of marks. Statistical moderation seeks to adjust for any arbitrary differences between schools in the standards of marking.

The method that the HKEAA uses to carry out statistical moderation follows well established international practice. In essence, the distribution of

SBA scores of students in a given school is made to resemble the distribution of scores of the same group of students on the public examination. The method adjusts the mean and the standard deviation of SBA scores, but the rank order of the SBA scores is not changed.

Results of statistical moderation

In the 2007 examination, 199 schools opted to submit SBA marks for feedback and to include the marks in the subject result, while 125 chose to submit SBA marks for feedback only. The mean and standard deviation of the SBA marks submitted by the majority of schools fell within the expected range.

Schools were given feedback in the form of an SBA Moderation Report in October 2007. In the report, two comments were given in addition to the mean and standard deviation of the SBA scores before and after moderation. The first comment related to the mean of the SBA scores awarded by teachers as a whole. If the school's SBA scores were within the expected range, only minimal adjustments were made. More adjustments were necessary for schools with means that were higher or lower than expected. The second comment was about the distribution of the SBA scores submitted by the school. If the standard deviation of the SBA scores was within the expected range, only slight adjustments were needed, while more adjustments were made to school scores with wider or narrower spreads than expected. A summary of the moderation results of Option 1 schools are given in Tables 5 and 6.

Table 5: Moderation results of the mean of SBA scores submitted by Option 1 schools

The mean of the SBA scores is ...	No. of Schools	Percentage (%)
within the expected range	144	72.4
slightly higher than expected	29	14.6
higher than expected	2	1.0
much higher than expected	0	0
slightly lower than expected	21	10.6
lower than expected	3	1.5
much lower than expected	0	0

Table 6: Moderation results of the S.D. of SBA scores submitted by Option 1 schools

The standard deviation of the SBA scores is ...	No. of Schools	Percentage (%)
as expected	179	89.9
slightly wider than expected	0	0
wider than expected	0	0
slightly narrower than expected	10	5.9
narrower than expected	9	4.5
much wider than expected	0	0
much narrower than expected	1	0.5

Since the SBA component carries a weighting of 15% of the public assessment, any upward or downward adjustment of the SBA marks has minimal impact on the overall subject result. For example, with a maximum of 48 marks for the English Language

SBA, an adjustment of 3 marks means a change of less than 1% to the subject total. Table 7 shows the impact of statistical moderation on the actual scores of the candidates.

Table 7: Moderation effect on candidates

Mark adjustment (% of subject mark)	No. of Schools	Percentage (%)
0 (0)	5365	17
1-3 (<1)	19881	62
4-6 (<2)	6237	20
7-9 (<3)	392	1
	31,875	100

The moderation results show that most teachers have a good understanding of the assessment criteria and can assess their students reliably. This is reassuring and indicates that the initial concerns about teacher readiness and fairness might have been exaggerated.

Analysis of 2007 examination and SBA data

Following the release of the 2007 HKCEE results, analyses of examination data for English Language were undertaken to determine:

- whether the different components of the exam measure a single underlying dimension;
- the reliability of each of the components; and
- the reliability of the composite score assuming 1) equal weights, 2) weights as set by HKEAA as a matter of policy, 3) weights that maximize the reliability of the composite score.

These questions were addressed by using structural equation modelling to fit a one-factor congeneric measures model to the data (Hill, 2007). The inter-paper correlations are given in Table 8.

Table 8: Inter-paper correlations (by listwise case exclusion, N=28,253)

	Reading	Writing	L & IS	Speaking	SBA
Reading	1.000	0.858	0.887	0.776	0.803
Writing	0.858	1.000	0.852	0.767	0.797
Listening & Integrated Skills	0.887	0.852	1.000	0.764	0.796
Speaking	0.776	0.767	0.764	1.000	0.787
SBA	0.803	0.797	0.796	0.787	1.000

The results of the analysis are summarized in Tables 9, 10 and 11 below. Table 9 indicates the extent to which the scores on the various parts of the examination measure a single underlying ability,

namely English Language. The table gives three ‘goodness-of-fit’ indices obtained from fitting a one-factor congeneric measures model to the data.

Table 9: Goodness - of - fit indices

Goodness of fit index	Adjusted goodness of fit index	Root mean square residual
0.969	0.908	0.017

The values indicate strong support for the existence of a single underlying ability for the various components of the English examination, including

SBA. This justifies the statistical moderation of SBA marks on the basis of the public examination scores.

Table 10: Reliability of the total scores weighted in different ways

Equal weights	Policy weights	Maximum reliability weights
0.957	0.959	0.961

The first column of Table 10 shows the reliability of the examination if equal weights are given to the various papers and to SBA. The second column shows the reliabilities of English Language given the weights that were actually assigned to the various components (e.g., Reading = 20%, Writing = 20%, etc.). It can be seen that with a reliability of 0.959, the English Language examination is highly reliable. The third column shows what the reliability would be if the various components were weighted in such a way as to maximize the reliability of the examination. It can be seen that the increase is very small relative to the actual policy

weights. This indicates that the weighting given to individual papers is in fact appropriate.

Table 11 provides information at the component level. In the first column are the relevant weights. In the second column are the factor loadings and in the third column, the variances of the residuals. Because correlation matrices were analyzed and variables were standardized, the reliability of each component, shown in the fourth column, is simply the square of the factor loadings. The factor score regressions in column five indicate the weights that one would use to maximize the reliability of the component.

Table 11: Reliability of the different components of the English Language examination

Component	Weight	λ_i	θ_i	Reliability of component	Factor score regressions
Reading	0.20	0.939	0.117	0.882	0.312
Writing	0.20	0.915	0.163	0.837	0.218
Listening and Integrated Skills	0.30	0.932	0.132	0.869	0.276
Speaking	0.15	0.838	0.297	0.702	0.110
SBA	0.15	0.868	0.247	0.753	0.137

It can be seen that all components were reliably measured and that the reliability of the SBA (which measures speaking) was higher than that for the speaking examination. This is contrary to most teachers' expectations, but should not come as a surprise. It is reasonable that multiple assessments conducted over the course of two years by students'

own teachers should be more reliable than a one-off 12-minute speaking examination taken under high-stress conditions.

From the above, it can be concluded that the CE English Language examination, including the SBA component, measured a single underlying ability and provided a highly reliable total score for each candidate

as well as reliable scores for each of the components of the examination. The initial doubts about the reliability of the SBA component can therefore be dispelled.

Effect of SBA on the speaking examination

Because of the SBA phase-in options offered to schools, the 2007 HKCE English Language Examination offers an opportunity for studying the effect of SBA implementation on the performance of candidates, in particular their speaking performance, since the SBA component also focuses on the assessment of speaking ability.

Absentee rate

The written papers for English Language are scheduled in early May each year while the speaking examination is conducted over a ten-day period in June, after the written papers for all other subjects have been sat. Candidates who have not done well in the written papers tend to give up on the speaking examination. Therefore, the absentee rate of the speaking examination has always been the highest among all English papers.

When a new examination syllabus is introduced, the absentee rate also tends to increase, possibly due to a lack of confidence on the part of candidates who may be unfamiliar with the new requirements. For example, in 1996, the last time when a major syllabus change was introduced to HKCEE English Language, the absentee rate in the speaking examination was 19.0%, up from 14.9% in 2006 and representing an increase of about 4%. The absentee rate eventually dropped to about 12.3% in 2006. In 2007, with the introduction of a new examination syllabus, the absentee rate for all candidates was 13.3%. There was an unexpectedly small increase of 1% as compared to the 2006 figures.

While the absentee rate of private candidates has remained fairly stable, at around 13%, the absentee rate of school candidates has fluctuated more markedly. Further analysis was done by dividing the schools into three groups based on their choice of SBA implementation option. A comparison of the absentee rates of different groups of school candidates is shown in Table 12.

Table 12: Absentee rate of candidates from different school groups

School choice	No. sat	No. of absentees	Absentee rate (%)
Option 1 (Yes)	27,935	3,398	10.8
Option 2 (Trial)	19,307	1,466	7.1
Options 1 & 2 (Yes + Trial)	47,242	4,864	9.3
Option 3 (No)	19,298	3,681	16.0
All Schools	66,540	8,545	11.4

The above figures reveal that the absentee rates of Option 1 and Option 2 schools are lower than that for Option 3 schools, which have deferred the implementation of SBA. A possible explanation for this difference is that candidates from Option 1 and Option 2 schools had more speaking practice in school because of the SBA. They were therefore more confident in taking part in the public speaking examination, which involves similar speaking skills required for the SBA tasks. It is also possible that the

candidates from Option 3 schools are generally weaker and therefore more prone to skip the speaking examination.

Speaking examination scores

A breakdown of the speaking examination scores of the candidates from different school groups reveals an interesting pattern, as shown in Table 13.

Table 13: Speaking examination scores of candidates from different school groups

School choice	No. sat	Speaking exam mean (%)
Option 1 (Yes)	27,804	25.38 (53)
Option 2 (Trial)	19,381	26.00 (54)
Option 3 (No)	21,293	23.51 (49)
All Schools	68,478	24.97 (52)

It can be seen that the mean speaking examination scores of candidates from Option 1 and Option 2 schools were higher than that of Option 3 schools. However, this cannot prove that the SBA component has a positive effect on the candidates' speaking performance. It could be argued that the schools opting for SBA implementation in 2007 were actually better schools with better students to begin with than those choosing not to implement SBA at all.

To further analyse the data, regression analysis was carried out to predict the speaking examination scores of candidates from different school groups

based on their scores in other English examination papers, which is taken as an indication of their general English ability. The actual and predicted scores were compared to see if there were any significant differences in the residuals (actual mean minus predicted mean). A positive residual would indicate that the group of candidates did better in the speaking examination than predicted based on their performance in other papers, which indicates their general English ability. Table 14 shows the differences between the actual and predicted mean scores.

Table 14: Actual and predicted speaking scores of candidates from different school groups

School choice	Actual Mean	Predicted Mean	Residual	t-value
Option 1 (Yes)	25.38	25.13	0.25	-1.97
Option 2 (Trial)	26.00	25.97	0.03	-0.21
Option 3 (No)	23.51	23.86	- 0.35	2.34 *

On average, candidates from Option 1 and Option 2 schools performed better in the speaking examination than expected, achieving higher mean scores than predicted although the differences are not statistically significant. It should also be noted that the residual of Option 1 schools, where the SBA scores were submitted and actually included in the subject results, is more positive than Option 2 schools, where SBA was implemented on a trial basis. However, candidates from Option 3 schools got statistically significant lower scores than expected, which means that they performed worse in the speaking examination relative to other papers.

It can be concluded that the implementation of SBA did have an effect on the performance of the candidates in the speaking examination. Candidates without SBA practice in school did significantly worse than expected, while those who participated in SBA did better than expected, regardless of their general English ability.

Chief Examiner’s comments

After each examination, the Chief Examiner of each paper submits a report which includes comments on the examination questions as well as on candidates’ performance. The following excerpts from the report on the speaking paper give an indication of examiners’ views on the effectiveness of the SBA:

“This year’s oral exam constituted a big change in format, but candidates were quite well prepared for this change. Also, thanks to continuous SBA practice, more students were more willing to contribute in both Parts A and B.”

“During the oral examination this year, the number of candidates who did not say anything at all dropped significantly. It was noticed that candidates were more confident and willing to talk... ..”

These observations are consistent with the statistical evidence. However, as it was not possible to distinguish between school candidates and private candidates in the examination room, or candidates who had or had not participated in SBA, these comments apply to all candidates who took the speaking examination. It would still be fair to say that there is anecdotal evidence that candidates’ speaking examination performance has improved in general after the introduction of the SBA.

Conclusion

The 2007 experience indicates that most teachers have a good understanding of the assessment criteria and can assess their students reliably in school. In the 2007 exam, 199 schools (34% of schools) chose

Option 1: to submit SBA marks for feedback and to include the marks in the subject result, while 125 (22%) of schools chose Option 2: to submit SBA marks for feedback only. The mean of the SBA marks submitted by 72% of the Option 1 schools and 65% of the Option 2 schools fell within the expected range, while 90% of the Option 1 schools and 94% of the Option 2 schools submitted marks with a spread within the expected range.

Statistically, the SBA component has proved to be as reliable as the public speaking examination. In fact, the moderated SBA marks correlated slightly better with the rest of the public examination papers than the speaking examination. This indicates that teachers are able to reliably assess the speaking abilities of their students given statistical moderation to remove arbitrary differences between schools in interpreting standards.

The beneficial washback effects of SBA can be seen in the results of the speaking examination. Candidates from schools that submitted SBA marks (Option 1 and Option 2 schools) had a lower absentee rate of about 9% compared to 16% for candidates from

schools that chose not to submit SBA marks (Option 3 schools). Candidates who did SBA in school also performed better in the speaking examination than those from schools not submitting SBA marks. Statistical evidence is supported by the Chief Examiner's comments and observations.

The 2007 experience has shown that the speaking ability of students can be reliably assessed in school by their own teachers and that the SBA is a valid and viable alternative to the speaking examination.

Further research is required to ascertain the validity and reliability of the SBA and its effect on the performance of candidates in the public examination. A four-year longitudinal study is being carried out to monitor the setting, conduct, marking and moderation of the SBA component of the HKCE English Language Examination over four school years (2005/06 to 2008/09). Analysis of the 2008 examination data is also underway and will provide more information on the effect of SBA on the performance of candidates as over 50% of schools have now chosen to implement SBA.

References

- Hill, P.W. (2007). *Reliability of 2007 HKCEE English Language*. Unpublished study.
- Hong Kong Examinations and Assessment Authority (2005). *2007 Hong Kong Certificate of Education Examination, English Language, Handbook for the School-based Assessment Component*.
- Hong Kong Examinations and Assessment Authority (2005). *Hong Kong Certificate of Education Examination, Regulations and Syllabuses 2007*.
- Hong Kong Examinations and Assessment Authority (2007). *2007 Hong Kong Certificate of Education Examination, English Language, Examination Report and Question Papers*.
- Hong Kong Examinations and Assessment Authority (2007). *Statistical Moderation of School-based Assessment Scores*.