School-Based Assessment in HKDSE:
Is the Statistical Moderation Method Disadvantageous for Students from Weak Schools?

WEN Zhonglin
(Hong Kong Examinations and Assessment Authority; zwen@hkeaa.edu.hk)

Abstract

In 2012, with the implementation of the new senior secondary academic structure, the Hong Kong Advanced Level Examination (HKALE) and the Hong Kong Certificate of Education Examination (HKCEE) were replaced by the Hong Kong Diploma of Secondary Education (HKDSE). School-Based Assessment (SBA) is an important component of the HKDSE, and one that enhances the validity of the assessment. A major concern regarding the implementation of SBA is whether the SBA marks submitted by schools, which are counted as part of students' public examination results, are fair and comparable across all schools. To address this concern, the Hong Kong Examinations and Assessment Authority has designed several moderation procedures to adjust the raw SBA marks submitted by different schools. These procedures include a statistical and non-statistical approach. As the statistical approach relies on the performance of the candidates' school in the public examination, there is a concern that the moderation method could disadvantage students from weak schools. However, the analysis of data taken from the 2013 HKDSE English Language subject shows that students from weak schools were not negatively affected by the SBA moderation, whatever grades they had, as the effect of being in a weak school was counteracted by the big-fish–little-pond effect.

Keywords: HKDSE, school-based assessment, moderation, big-fish–little-pond effect, reflected-glory effect

The Hong Kong Diploma of Secondary Education (HKDSE) is an academic qualification offered by the Hong Kong Examinations and Assessment Authority (HKEAA). It is the only public examination in the new 3-3-4 education system introduced in Hong Kong secondary schools, and replaces the HKCEE and HKALE as the new benchmark examination for the Joint University Programmes Admissions System (JUPAS, the undergraduate admission allocation system in Hong Kong. For more details please refer to http://www.jupas.edu.hk/en). Most school candidates take four core subjects in the HKDSE (Chinese Language, English Language, Mathematics and Liberal Studies), plus two or three elective subjects.

School-based Assessment (SBA) is a salient feature of HKDSE. In this paper, we briefly describe the development of SBA and SBA moderation in Hong Kong, and investigate whether students from weak schools are disadvantaged by the SBA moderation procedure.

The development of SBA in Hong Kong

In the context of public assessment, SBA refers to assessments administered in schools and marked by the students' own teachers. SBA marks awarded count towards students' public assessment results. SBA has been practised in Hong Kong for more than 30 years.

SBA in HKALE

As early as 1978, SBA was introduced as part of the Hong Kong Advanced Level Examination (HKALE), the entrance examination for universities in Hong Kong until 2012. The HKALE has now been replaced by the HKDSE examination as part of local education reforms.

SBA was implemented in 14 HKALE Advanced (A) and/or Advanced supplementary (AS) subjects (see Table 1). The first of these was A Chemistry, with SBA implemented from 1978, and the last was A Computer Studies, with SBA implemented from 2007. The minimum weighting of SBA was 10% (AS Chinese Language & Culture), and the maximum weighting was 35% (AS Visual Arts). The medium weighting of SBA was 20%.

Table 1. SBA in HKALE Subjects

Subjects	SBA implemented from	Weighting
A Chemistry	1978	20%
A Government & Public Affairs	1988	22.5%
AS Chemistry	1994	20%
AS Chinese Language & Culture	1994	10%
AS Liberal Studies	1994	20%
A Biology	1995	20%
AS Electronics	1999	20%
AS Computer Applications	2000	30%
A/AS Physics	2004	15%
A/AS Art	2004	25%/35%
A Chinese Literature	2005	25%
A Computer Studies	2007	20%

(Source: http://www.hkeaa.edu.hk/en/SBA/SBA_HKALE_bg.html)

SBA in HKCEE

SBA was also conducted as part of the Hong Kong Certificate of Education Examination (HKCEE), a public examination for students who had completed a full-time five-year secondary school course, until 2011. The HKCEE was discontinued in 2012 and its roles are now being replaced by the HKDSE examination.

SBA was implemented in 14 HKCEE subjects (see Table 2). The first of these were Design & Technology, and Electronics & Electricity, with SBA implemented from 1980, and the last were Chinese Language and English Language, with SBA implemented from 2007. The minimum weighting of SBA was 15% (Chinese Language, English Language), and the maximum weighting was 50% (Visual Arts). The medium weighting of SBA was 25%.

Table 2. SBA in HKCEE Subjects

Subjects	SBA implemented from	Weighting
Design & Technology	1980	33.3%
Electronics & Electricity	1980	30%
Fashion & Clothing	1989	30%
Design & Technology (Alt. Syl.)	2002	30%
Graphic Communication	2002	30%
Technological Studies	2002	30%
Computer & Information Technology	2005	20%
Integrated Humanities	2005	20%
Science & Technology	2005	20%
Visual Arts	2005	50%
Chinese History	2006	20%
History	2006	20%
Chinese Language	2007	15%
English Language	2007	15%

(Source: http://www.hkeaa.edu.hk/en/SBA/SBA_HKCEE_bg.html)

SBA in the HKDSE Examination

Since 2012, the first year of the HKDSE examination, a total of 13 subjects (3 core subjects and 10 elective subjects) have had SBA as a component. The minimum weighting of SBA is 15% (English Language), and the maximum weighting is 50% (Visual Arts). The medium weighting of SBA is 20%. Details are listed in Table 3.

Table 3. SBA in HKDSE Subjects

Subjects	Weighting
Chinese Language	20%
English Language	15%
Liberal Studies	20%
Chinese History	20%
Design & Applied Technology	40%
History	20%
Information & Communication Technology	20%
Visual Arts	50%
Biology	20%
Chemistry	20%
Physics	20%
Integrated Science	20%
Combined Science	20%

(Source: http://www.hkeaa.edu.hk/en/SBA/sba_hkdse/SBA_timetable.html)

Why SBA?

The primary purpose of SBA is to enhance the validity of the assessment by including the assessment of outcomes that cannot be readily assessed within the context of a one-off public examination. SBA can also reduce dependence on the results of public examinations, which may not always provide the most reliable indication of candidates' actual abilities. Obtaining assessments that are developed by those who know the students best – their subject teachers – and that are based on student performance over an extended period of time, provides a more reliable picture of each student's abilities. SBA also has a positive impact on teaching and learning: helping to motivate students by engaging them in meaningful activities, reinforcing curriculum aims and good teaching practices among teachers, and providing structure and significance for daily teaching activities, namely, for teachers to assess their own students (http://www.hkeaa.edu.hk/en/sba/).

Taking English Language as an example, the rationale laid down by the SBA Consultancy Team (2005) for implementing SBA was as follows:

- to continuously assess students in a pressure-free environment;
- to reduce reliance on 'one-off' public oral examination;
- to improve the reliability of oral English assessment;
- to reflect the standard and ability of students;
- to foster teaching and learning;
- to promote students' leisure reading and listening;
- to reinforce learners' autonomy and independent learning;
- to facilitate 'learning how to learn' through peer reviews;
- to inform prospective employers and universities of students' level of ability;
- to make Hong Kong's examination system in line with the international model so that 'assessment for learning' is achieved; and
- to empower teachers to make part of the assessment mechanism.

Why Moderation?

One important concern regarding the implementation of SBA is the fairness and comparability of the SBA marks submitted by schools. Teachers know their students very well and are thus 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 judgements, they are not necessarily aware of the standards of performance across other schools. Despite the fact that teachers are given training in how to carry out SBA, and assess students on the same task using the same assessment criteria, teachers in one school may be harsher or more lenient in their judgements than teachers in other schools, or may tend to use a narrower or wider range of marks.

To address these potential problems, the HKEAA employs appropriate methods for 'moderating' the raw SBA scores submitted by different schools, with the following aims (HKEAA, 2011):

- to maintain comparability of SBA results across schools, and thus ensure fairness for individual students and schools;
- to maintain the quality, reliability, and validity of SBA from year to year; and
- to gather information that may be useful for making recommendations for improved practice (feedback to schools).

Statistical moderation

Statistical moderation is particularly appropriate in situations in which another measure is available that can be used to 'moderate' schools' assessments. Typically this other measure will be students' performance in the examination in that subject or on a test of general academic ability. The main advantages of statistical moderation are that it can be carried out quickly, impartially and without a large commitment of time and resources. Although the method is reliant on the assumption that the measure used to moderate assessments is a valid measure of the overall level of the performance of students in the moderating group, this assumption is generally valid in the context of most academic subjects subject to public examination.

There are essentially two ways in which differences in marking standards may affect SBA scores. First, teachers in a given school may be either harsher or more lenient than teachers in other schools. Second, they may tend to either overly bunch students' scores together or spread them apart too much.

The statistical moderation method designed by the HKEAA addresses both of these potential problems by adjusting the average and the spread of SBA scores of students in a given school so that they are in alignment with the examination scores of the same group of students. What this means in practice is that the average or mean of the SBA scores may be moved up or down, and the spread or standard deviation of scores stretched or compressed, in line with the mean and standard deviation of the examination scores of the same group of students. While this ensures that SBA scores are comparable across schools, it does not, however, change the rank ordering of SBA scores within a school. More detailed information on the statistical moderation method can be found on the HKEAA website (HKEAA, 2011).

Does the statistical moderation method disadvantage students from weak schools?

As statistical moderation relies on the performance of the moderation group (usually one school as one group) in the public examination, there is a concern as to whether the students

from weak schools will be put at a disadvantage. In response to this concern, we conducted a study using live data from the 2013 HKDSE English Language subject.

In this study, schools were classified into three bands according to their performance in the English Language public examination. The schools with better performance were Band 1, the schools with medium performance were Band 2, and the schools with poorer performance were Band 3. All candidates were grouped by: (1) their subject grades for English Language (grade group); and (2) their percentage rank in the English Language public examination (percentage group). In each grade group or percentage group, differences among bands were compared and analysed.

As only a few candidates from Band 1 schools were ungraded, it was not statistically meaningful to conduct difference analysis among ungraded candidates. These candidates were therefore excluded from further analysis.

Differences among Bands by Grade Group

Table 4 lists the results of the comparison of moderated SBA for English Language among different bands for candidates in the same grade group (see also Figure 1). The results show that the means of moderated SBA increased as grades rose. For each grade, the differences of the means of moderated SBA among bands were not larger than 3.0 (the full mark was 100), although the differences were statistically significant. Given that the weight of SBA in English Language was 15%, a difference of 3.0 in SBA marks results in a trivial mark difference of 0.45% in the total subject mark. The band effect was ignorable as the largest Eta was 0.07 and its square[1] (effect size) was 0.00.

Table 4. Comparison of the means of moderated SBA scores for English Language among different bands by grade group

Grade	Band	N	Mean	SD	Max–Min	Sig.	Eta
1	1	110	44.1	9.60	2.7	.00	.04
	2	3962	41.4	7.76			
	3	5282	41.4	8.42			
2	1	2888	55.3	7.25	1.7	.00	.07
	2	12540	53.7	7.64			
	3	4834	53.9	8.50			
3	1	8859	65.2	6.64	1.1	.00	.04
	2	6192	64.9	7.15			
	3	1046	66.0	7.98			
4	1	9367	73.9	6.45	0.5	.02	.03
	2	1566	74.4	7.33			
	3	170	74.0	7.43			
5	1	6392	84.5	7.18	3.0	.02	.04
	2	254	83.6	7.08			
	3	21	81.5	7.05			

---

[1] The squared Eta is the most frequently reported effect size in ANOVA. The squared Eta=SS_effect/ SS_total.

The results showed that there were no substantial differences among the means of moderated SBA across three bands for any grade. No band was put at an advantage or disadvantage under the SBA moderation procedure. Hence, for English Language, the SBA moderation procedure appears fair.
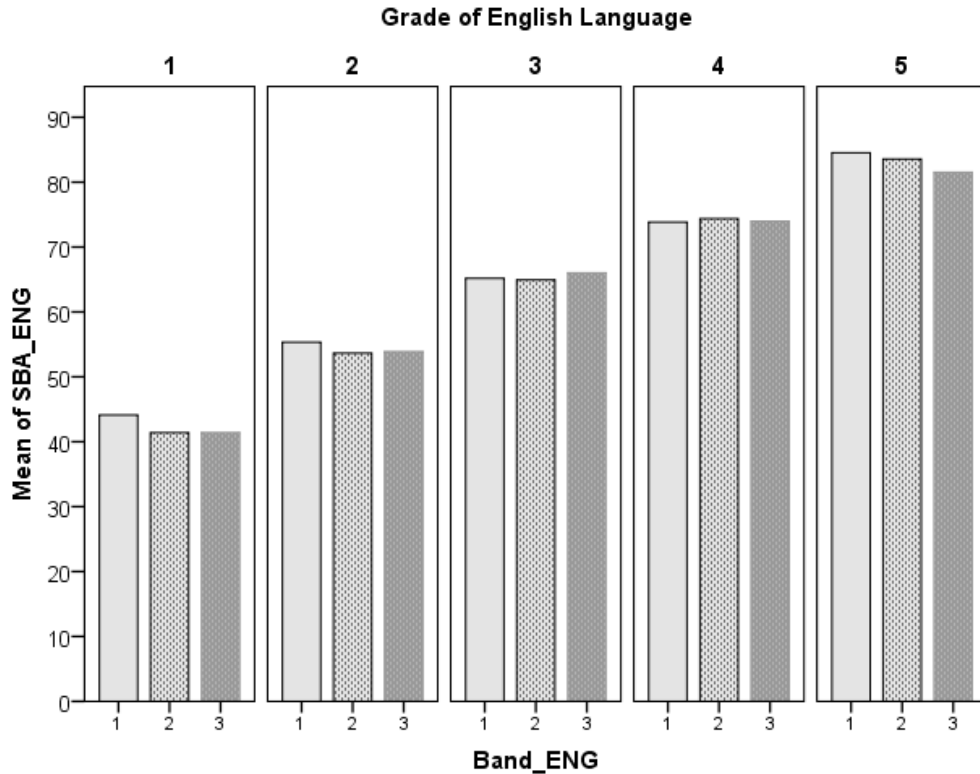

Figure 1. Comparison of the means of moderated SBA scores for English Language among three bands

Difference among Bands by Percentage Group

Table 5 lists the results of the comparison of moderated SBA for English Language among different bands for candidates in the same percentage group (see also Figure 2). As there were only seven candidates from Band 1 schools in the group of "0–10%", it was not statistically meaningful to conduct difference analysis in this group.

For the groups from "20–30%" to "80–90%", the differences of the means of moderated SBA among bands were less than 3.0, leading to a trivial mark difference of less than 0.45% in the total subject mark. The band effect was ignorable as the largest Eta was 0.08 and its square (effect size) was 0.01.

For the group of "10–20%", the mean of moderated SBA of Band 1 was almost 4 marks higher than those of the other two bands (corresponding to 0.6% of the total subject mark), indicating the reflected-glory effect (Marsh, Kong, & Hau, 2000). That is, poor students in Band 1 schools were likely to obtain a slightly higher SBA mark than poor students in Band 2 or Band 3 schools.

For the group of "90–100%", the mean of moderated SBA of Band 3 schools was almost 4 marks less than that of Band 1 schools, and the (reverse) reflected-glory effect occurred in this

group. That is, excellent students in Band 3 schools were likely to obtain a slightly lower SBA mark than excellent students in Band 1 schools.

The results show that, for the majority of students, there were no substantial differences in the means of moderated SBA among the three bands. Although excellent students in Band 3 schools were likely to be at a slight disadvantage, the level of disadvantage was minor and even ignorable.

Table 5. Comparison of the means of moderated SBA scores for English Language among different bands by percentage group

| Percentage	Band	N	Mean	SD	Max–Min	Sig.	Eta
0–10%	1	7	52.7	12.70	11.4	.00	.14
	2	234	41.4	7.38			
	3	741	41.5	6.14			
10–20%	1	74	45.7	10.54	3.9	.00	.05
	2	2988	41.9	7.94			
	3	4074	41.8	8.71			
20–30%	1	290	49.0	7.75	0.8	.00	.05
	2	4182	48.3	7.63			
	3	2591	49.1	8.78			
30–40%	1	737	53.5	7.32	1.5	.00	.08
	2	4473	53.1	7.59			
	3	1592	54.6	8.60			
40–50%	1	1559	57.8	6.84	1.1	.00	.05
	2	4107	57.6	7.61			
	3	1049	58.6	9.42			
50–60%	1	2709	61.6	6.83	1.4	.00	.06
	2	3354	61.6	7.30			
	3	660	63.0	9.35			
60–70%	1	3904	65.6	6.92	0.8	.03	.03
	2	2462	66.0	7.57			
	3	358	66.4	9.51			
70–80%	1	5113	69.6	6.82	0.9	.27	.02
	2	1555	69.5	8.06			
	3	188	70.4	8.64			
80–90%	1	6171	74.6	6.86	2.4	.02	.03
	2	860	74.8	7.76			
	3	73	72.4	12.32			
90–100%	1	7052	83.3	7.83	3.8	.00	.05
	2	299	81.7	8.18			
	3	27	79.4	7.56			
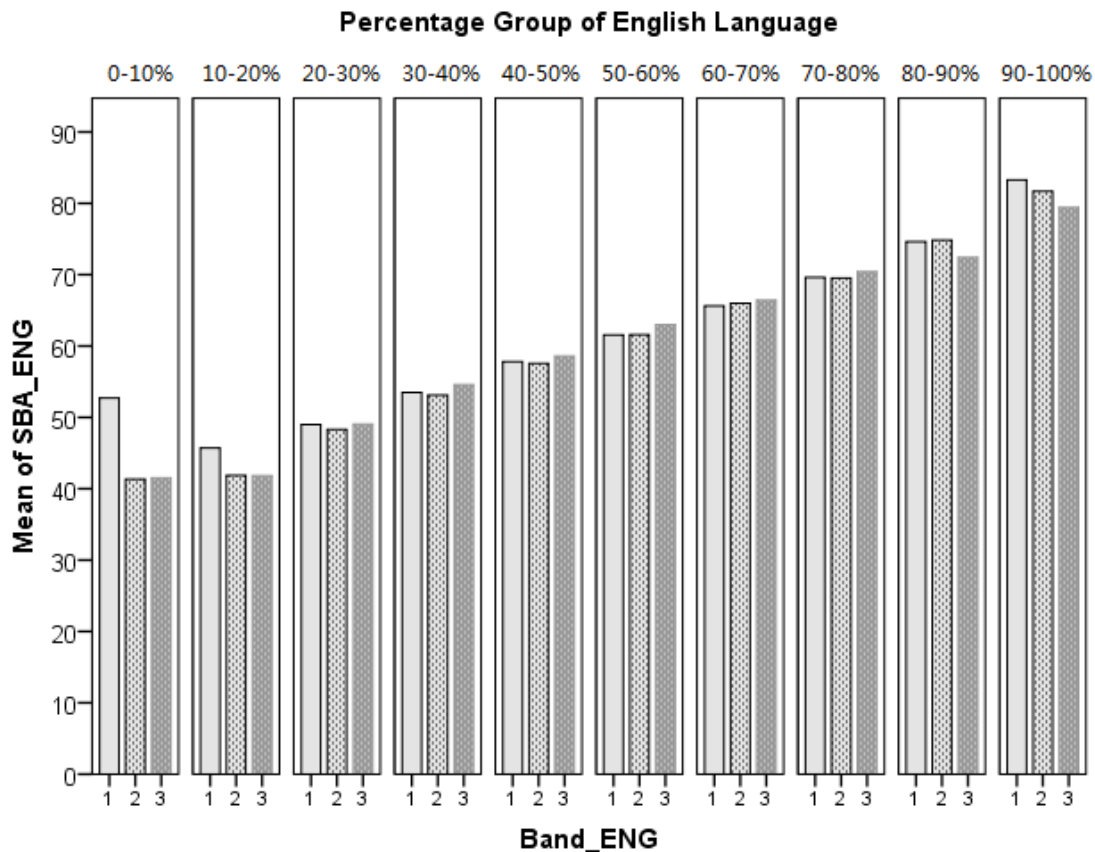
Figure 2. Comparison of the means of moderated SBA scores for English Language among three bands (candidates were grouped by percentage rank in public examination)

Conclusion and discussion

Candidates' raw SBA marks were moderated according to two key factors: the first being their school's performance level in the public examination, the second their individual SBA performance as marked by the subject teacher (i.e., their raw SBA mark). On one hand, competent students in Band 3 schools were put at a disadvantage because of the lower performance level of their school in the public examination (reflected-glory effect). On the other hand, the students gained an advantage because of the big-fish–little-pond effect (BFLPE, Ludtke, Koller, Marsh, & Trautwein, 2005; Marsh & Hau, 2003). That is, they tended to be given a higher SBA mark (raw data) because they were outstanding when compared to their relatively weak schoolmates.

Thus, the SBA moderation results were influenced by the reflected-glory effect and the big-fish–little-pond effect (BFLPE). The combination of these two effects generally resulted in a counterbalance and a fair SBA moderation procedure. When the BFLPE was larger than the reflected-glory effect, competent students in Band 3 were likely to gain a slight advantage after SBA moderation. When the reflected-glory effect was larger than the BFLPE, competent students in Band 3 were likely to be put at a slight disadvantage.

For the 2013 HKDSE English Language, the SBA moderation was fair as far as grade groups were concerned because there was no discernible band effect for each grade group. However, when percentage groups were considered, the exceptions were at the top and bottom percentage groups. In the top percentage group, the BFLPE was not sufficient to counteract the demonstrated (reverse) reflected-glory effect because students from Band 3 seemed to have been slightly disadvantaged. At the opposite end, in the bottom two groups, students from Band 1 might have enjoyed reflected-glory effect. Nevertheless, the difference of 4% in SBA would only lead to a difference of 0.6% of the total subject mark. It was really negligible compared to other errors. In the bottom percentage group, poor students in Band 1 schools obtained substantial higher marks (11% in SBA), but since the number of students involved was only 7, the possible effect was not statistically meaningful and more research is needed in this area.

Reference

Education and Manpower Bureau. (2005). *The new academic structure for senior secondary education and higher education – Action plan for investing in the future of Hong Kong*.

Hong Kong Examinations and Assessment Authority (2011). *Moderation of School-based Assessment Scores in the HKDSE*.
http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/HKDSE-SBA-ModerationBooklet_r.pdf

Ludtke, O., Koller, O., Marsh, H.W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, *30*, 263–285.

Marsh, H., W., & Hau, K.-T. (2003). Big-Fish–Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist, 58*, 364–376.

Marsh, H., W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish–little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78,* 337–349.

SBA Consultancy Team. (2005*). 2007 HKCEE English examination: Introduction to the school-based assessment component*. Hong Kong: HKEAA. http://en.wikipedia.org/wiki/School-based_assessment