Review of Moderation Method for Independent Enquiry Study (IES) Raw Marks in 2012 HKDSE Liberal Studies

Or Chuen-san, Albert
aor@hkeaa.edu.hk
Hong Kong Examinations and Assessment Authority

Fung Tze-ho, Eric
thfung@hkeaa.edu.hk
Hong Kong Examinations and Assessment Authority

Abstract

In the assessment framework of 2012 HKDSE Liberal Studies (LS), school-based assessment (SBA) is an important component, in addition to written exam. The mode of SBA in LS is Independent Enquiry Study (IES), where each candidate is required to complete an enquiry on a current social issue. Candidates need to adopt suitable methods of data collection and enquiry. For the sake of performance standard comparability across schools, the Hong Kong Examinations and Assessment Authority (HKEAA) has to moderate IES raw marks submitted by schools. In view of specific nature of IES, a high correlation between exam results and IES performance may not be guaranteed. Thus, a hybrid moderation method is needed with using both schools' written exam results and re-assessment results on samples collected from schools.

After the completion of 2012 HKDSE, a study was conducted to examine the hybrid moderation method based on a large scale re-marking of sampled student works by experienced assessors. It is found that the moderation results from the hybrid method were closer to the 'true' values of school IES performance, than the one using written exam results alone. It proves that the hybrid moderation method to the nature of IES in LS is necessary.

Keywords: School-based Assessment, Statistical Moderation, Sample Re-assessment

Introduction

SBA is a salient feature of the HKDSE Examination. SBA refers to assessments administered in schools and marked by the students' own teachers. SBA in LS requires each student to carry out an Independent Enquiry Study (IES). The IES provides a valuable opportunity for students to independently carry out a focused enquiry into a current social issue of interest, and to present their views, ideas, findings, evaluations and personal reflections.

The raw SBA marks are submitted from schools to HKEAA. After receiving the raw SBA marks, the Authority has to undertake the SBA moderation process. The main reason for carrying out moderation is to ensure the consistency of assessment standards across schools. Teachers know their students well and thus are best placed to judge their students' relative performance. However, they could not be aware of the standards of performance across all schools. Therefore, teachers in some schools may be harsher or more lenient in their judgment than teachers in other schools. Mark ranges of scores awarded in various schools may also be different from each other.

To resolve these problems, the HKEAA employs appropriate methods for 'moderating' the raw SBA scores submitted by different schools to achieve the following:
(i)   The comparability of SBA results across schools in order to ensure fairness for individual students and schools
(ii)   The quality, reliability and validity of SBA results
(iii)   Provision of useful feedback to schools for improving practice

In LS, the SBA moderation is conducted using a hybrid method. The method is based on statistical moderation results using schools' written exam performance and re-assessment results on samples collected from schools. In the following, we briefly introduce statistical moderation and sample re-assessment, explain how to integrate these results together.

Statistical Moderation

Statistical moderation is particularly appropriate in situations where there is another measure available that can reflect SBA performance level. Typically this other measure will be students' performance in the public examination of that subject. An advantage of the method is that it can be carried out efficiently and impartially within a reasonable amount of time and resources. The key assumption is that the overall performance in the public examination of students in a school can properly reflect the SBA performance level of the same group of

students. Generally speaking, this is a valid assumption in the context of many academic subjects in public examinations.

In the moderation process, the adjustments are applied only to school average and spread of raw SBA scores of students with reference to their public examination scores in the same school. Therefore, the ranking of students within a school remains unchanged after moderation. The school averages of examination scores are used to determine the corresponding performance levels on SBA, taking within-school correlations between students' raw SBA scores and examination scores into consideration (HKEAA, 2010).

Sample Re-assessment

Some of the objectives of the SBA cannot be precisely assessed in the public examination due to different requirements. Moreover, students in SBA would possibly gain significant improvement under teachers' supervision due to the efficacy of assessment for learning. If only schools' public examination scores are used to adjust students' raw SBA scores, for some schools the statistical moderated results may not fully reflect the students' actual performance in the SBA; i.e., there may be some outlier schools whose statistically moderated scores differ greatly from the performance level demonstrated by students' SBA work. Therefore, for 2012 HKDSE LS, each school was required to submit six samples of students' work for reassessment which was conducted by a group of external assessors appointed by the Authority. The samples were chosen by the Authority using stratified random sampling. Students in each school were divided into six strata based on their raw SBA scores. Therefore, in each stratum the performance level of students on SBA should be similar with each other. Some students' work was then randomly chosen from each stratum. The stratified sampling method could ensure that a fairly small sample of students' work could adequately represent the full range of SBA performance of each school.

All the LS samples were then reassessed with reference to the previous standardized exemplars and a set of stipulated assessment criteria. If prominent discrepancies between external assessors' scores and raw scores were observed, discrepancy marking would be conducted. It was observed that the discrepancy marking percentage was about 20% in 2012. The correlation between raw scores and results based on external assessment amounted to 0.8. This reflected that the marking standards of school teachers were generally in line with that based on external assessment.

With regard to possible sampling variations, the ratio of school average of raw SBA scores to

sample average of raw SBA scores was examined for each school. The distribution of these ratios of 523 schools is shown below.

Figure 1: Distribution of ratios of school means of raw SBA scores to sample means of raw SBA scores
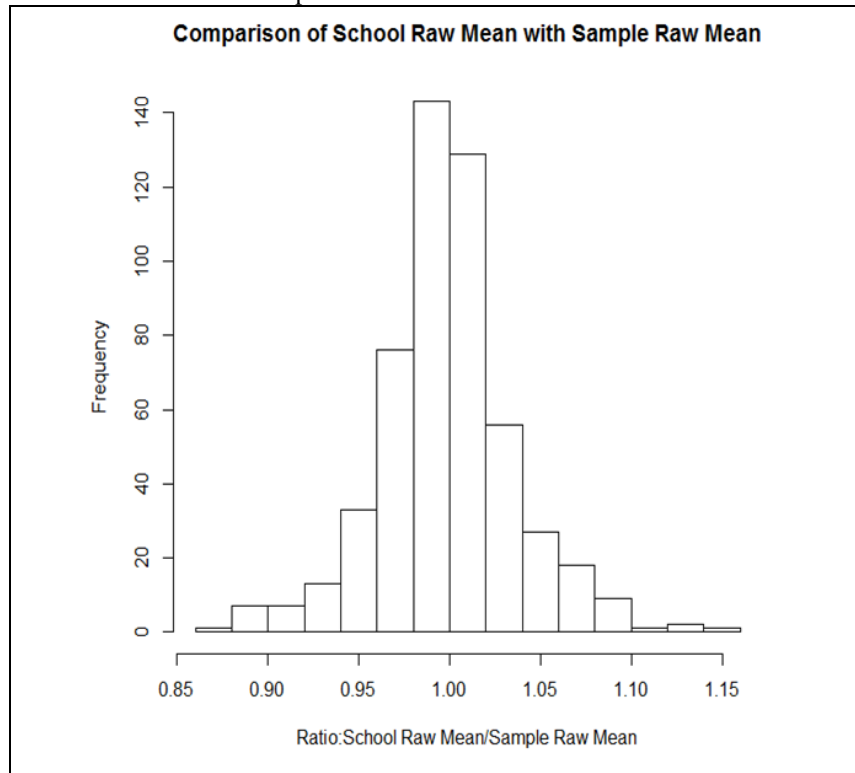
The 5th percentile of the distribution was 0.94 and the 95th percentile was 1.06. It implied that sample raw means were very close to school raw means for most schools. In addition, it should be noted that the mean mark of sampled students' work from external assessment of a school would be adjusted upwards when sample raw mean was less than school raw mean; and vice versa. With such adjustments, it was expected that the sampling variations would be further reduced.

To further enhance the reliability of the estimations of means and spreads of SBA scores of schools based on external assessment, Bayesian hierarchical modeling was employed so as to share information across different schools. The model is briefly described below.

Let $\mathbf{Y_i}$ (a vector) be the marks based on external assessment of a school i; i.e., $Y_{i,1}$, $Y_{i,2}, Y_{i,3},...,Y_{i,ni}$. The number of students in the school is n_i. The Bayesian hierarchical model is set up as follows:

$Y_{i,1}, Y_{i,2}, Y_{i,3},...,Y_{i,ni} \sim \text{Normal}(\theta_i, \sigma_i^2)$

for i = 1,…,m (i.e., there are m schools)

$\theta_i \sim$ Normal(μ, τ^2)

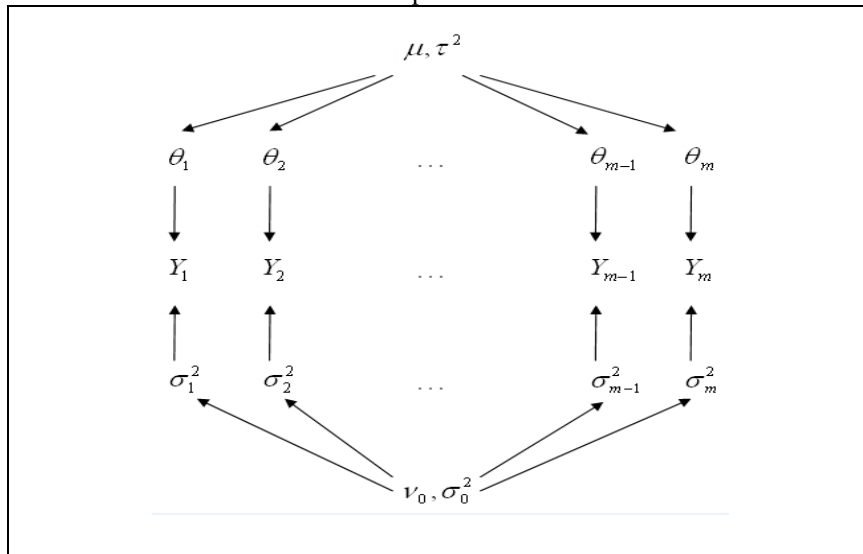for i = 1,$\cdots$,m (i.e., all i are sampled from a super-population)

$1/\sigma_i^2 \sim$ Gamma($v_0/2$, $v_0\sigma_0^2/2$)

for i = 1,…,m (i.e., all σ_i^2 are sampled from a super-population)

The model is graphically displayed in the figure below.

Figure 2: The structure of Bayesian hierarchical modeling showing relationship between data observed and parameters involved

In Bayesian analysis, the parameters: μ, τ^2, v_0, and σ_0^2 are treated as random variables. To conduct the Bayesian estimation, some non-informative priors p(μ), p(τ^2), p(v_0), p(σ_0^2) are set up respectively for μ, τ^2, v_0, and σ_0^2. Based on such a model, information could be shared across schools when estimating θ_i and σ_i^2. For schools extreme empirical values, the estimates of θ_i and σ_i^2 will be pulled towards the corresponding overall estimates (μ and σ_0^2). The θ_i values would be used for reference to further adjust school SBA performance results obtained from sample re-assessment.

In general, algorithms using Markov Chain Monte Carlo (MCMC) method are deployed for estimation in Bayesian hierarchical modeling. It is well known that such a hierarchical model could reduce the estimation error (James O. Berger, 1993, Gelman *et. al.*, 2003, Peter D. Hoff, 2010) in different applications. In addition, a simulation study has been undertaken to gauge the magnitude of gain in accuracy when applying the model in the specific setting for SBA moderation (Fung, 2011). It is found that the total Mean Squared Error (MSE) in the estimation of school means could be reduced by some 30% using Bayesian hierarchical modeling, as compared with the one simply using sample means.

Hybrid Method

After consolidating the sample review result of a school, it was compared with the corresponding result from statistical moderation. Due to possible variations incurred in the sampling and remarking process, an appropriate tolerance limit was set when comparing the two results. If the difference was within the tolerance limit, the statistical moderation result would be adopted as the school performance level on SBA. If the difference exceeded the tolerance limit, appropriate adjustments would be made to the statistical moderation result with reference to the sample review result in order to determine the school performance level on SBA.

Review of Moderation Results

After the completion of 2012 HKDSE, a study was conducted to examine the hybrid moderation method based on a large scale re-marking of sampled student works by experienced assessors. 10 schools covering a wide spectrum of IES performance levels are selected. From each school, some 50 student projects are sampled using stratified sampling based on raw marks for re-assessment by external assessors.

After collecting some 50 student projects from a school, the 'true' school performance level is derived by re-assessing the sampled projects. The 'true' school performance levels are then compared with the results obtained from statistical moderation alone and that obtained from hybrid method.

Profile of Schools Selected

Totally 10 schools were invited to join the study based on the expertise and recommendation of colleagues in Assessment Development Division. The medium of instruction (MOI), the performance level in past public exam and raw means of Task marks in 2012 HKDSE LS of these schools are tabulated below.

Table 1: School profile in terms of MOI, performance level in past public exam and
raw mean of project mark**s**

School Code	MOI	Performance level in past public exam	Raw mean of Project marks
20229	CMI	Middle	30.7
20230	CMI	High	33.7
20364	CMI	High	27.7
20371	EMI	High	32.1
30614	CMI	Low	19.5
30620	CMI	High	33.6
30758	CMI	Middle	28.9
30762	CMI	Middle	28.5
30825	CMI	Low	22.5
30790	CMI	High	39.9

From Table 1, it can be observed that there are five schools with high level of performance, three schools with middle level, and two schools with low level. Most of their media of instruction are Chinese. Only one of them uses English as MOI. The spread of school raw means of these 10 schools are quite wide, ranging from 19.5 to 39.9.

Study Results

We examined the moderation results from Statistical Moderation based on exam results and Sample Re-assessment results with the 'true' school performance levels on SBA which were determined by re-marking some 50 samples from each school. Similar to the practice in 2012 HKDSE LS exam, discrepancy marking was undertaken in the re-marking exercise of student projects, when there was a large discrepancy between raw mark and re-assessment mark (i.e., more than or equal to three marks out of nine). In this part of the study, 10 schools were involved and the comparison results are tabulated below.

Table 2: Comparison of school SBA means from Statistical Moderation and from Sample Re-assessment with the 'True' ones

School Code	Count	True SBA Level	Stat Mod	Sample Re-assessment
20229	53	28.0	28.8	28.7
20230	66	32.9	31.7	33.9
20364	57	29.2	30.4	28.2
20371	64	32.1	31.2	31.5
30614	48	18.0	26.0	22.2
30620	63	34.0	31.1	32.0
30758	42	27.7	29.3	30.5
30762	50	27.6	28.1	31.1
30825	50	23.3	29.3	26.1
30790	52	38.9	31.3	39.7
Std Dev	-	5.8	1.8	4.7

From Table 2, the following can be observed:

(i) The schools 30614 and 30825 have the prominent low levels of performance on SBA amongst the schools. Such situations could not be well reflected in the results from statistical moderation. On the other hand, the results from sample re-assessment could capture these two cases better.

(ii) The school 30790 has the highest level. However, such a situation could not be well reflected in the results from statistical moderation. On the other hand, the results from sample re-assessment could capture the extreme case better.

(iii) Such a phenomenon is also illustrated by examining the corresponding standard deviations of school means. The standard deviation of 'true' means is 5.8, and the standard deviation of results from sample re-assessment is 4.7; while that from statistical moderation only amounts to 1.8.

(iv) In general, statistical moderation results could capture quite well the school SBA performance level, except for the schools with extremely good and bad performance. In these cases, results from sample re-assessment based on only six samples selected from a school using stratified sampling could be a good reference for determining school SBA performance level.

References

Fung, T. H. (2011). Simulation Study on the Use of Hierarchical Bayesian Modeling in Expert Judgment for SBA Moderation. *International Meeting of the Psychometric Society*.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*, (2nd edn). London: Chapman & Hall/CRC.

Hong Kong Examinations and Assessment Authority (HKEAA). (2010). *Moderation of School-based Assessment Scores in the HKDSE Booklet*. Available from:http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE-SBA-ModerationBooklet_r.pdf

James O. Berger (1993). *Statistical Decision Theory and Bayesian Analysis*, (2nd edn). Springer.

Peter D. Hoff (2010). *A First Course in Bayesian Statistical Methods*, (1st edn). Springer.