Critique of the Research Article "Views from the Chalkface: English Language School-Based Assessment in Hong Kong" by the Researcher "Chris Davison" (2007)

Hastings Chim Ho Yeung[1,*]

[1]University of Bristol, UK

*Correspondence: University of Bristol, UK. E-mail: hastings.chim@bristol.ac.uk

Abstract

This essay critically analyzes "Views from the chalkface: English language school-based assessment in Hong Kong," a published study by Davison (2007). The conceptual framework of this essay draws on Stanley, MacCann, Gardner, Reynolds, and Wild's (2009) "Review of teacher assessment: Evidence of what works best and issues for development" and Cheng, Andrews, and Yu's (2011) "Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong." "Review" is used to critique Davison's methods and claims, whereas "Impact" is used to evaluate his findings. The essay ends with the writer's experiences in both monitoring and implementing SBA while he was serving as the English panel head of a band-three secondary school.

Keywords: School-Based Assessment, HKCEE, Assessment evaluation

1. Introduction

This essay will primarily focus on analyzing Davison (2007) article "Views from the chalkface: English language school-based assessment in Hong Kong", on the basis on the conceptual framework drawn on two works from Stanley, MacCann, Gardner, Reynolds, and Wild's (2009) and Cheng, Andrews, and Yu's (2011). The first paper will be utilized to evaluate Davison's claims and methods whilst the second paper is adopted to review his findings. A conclusion including the writer's frontline experience in conducting SBA will be given.

2. Summary of Davison's 2007 study

Davison's article was published when the Hong Kong Examinations and Assessment Authority (HKEAA) transformed the university entry examination, namely, the Hong Kong Certificate of Education Examination (HKCEE), from norm-referenced to standard-referenced. This transformation introduced a school-based oral assessment component, namely, School-Based Assessment (SBA), to the classroom. Such an unprecedented transformation of the assessment framework aroused concern among various stakeholders with regard to sociocultural, technical, and practical issues, given that SBA was the first step of the unification of assessment and curriculum reform by the Education and Manpower Bureau. Davison's article focuses on these issues by outlining the rationale and framework of the SBA component, reviewing analyses of how teachers and students view SBA during the pilot stage of SBA implementation, and proposing possible research on fostering the skills and fundamental knowledge of teachers in relation to SBA.

3. Stanley, MacCann, Gardner, Reynolds, and Wild's study as theoretical framework

Davison (2007) suggested that the emergence of SBA can be classified into three separate directions, namely, sociocultural, technical, and practical issues, with the new assessment skeleton. Therefore, Stanley et al.'s "Review," which traces how assessment innovation has been reformed globally, can be used to dissect Davison's spectrum and determine how the practical and theoretical perspectives in a globalized arena can be streamed to a localized one.

4. Limitations of Stanley, MacCann, Gardner, Reynolds, and Wild's study

Despite the resemblance of their perspective to that of the writer, Stanley, MacCann, Gardner, Reynolds, and Wild comprehensively analyze the factors influencing assessment innovation and its implementation in a global setting. As a result, their focus on the school-based spoken test is not specifically convergent. Their analysis is narrowed to compare the distinctive features derived from global and local assessment practice, instead of examining in depth all features of the assessment of oral proficiency.

5. Cheng, Andrews, and Yu's study as theoretical framework

Davison observed that students and parents generally had positive attitudes toward SBA implementation aside from teacher training and resources. Seven years has passed since 2007. Thus, Cheng, Andrews, and Yu's investigation of the opinions of parents and students can best be served by a retrospective critical review of the assumptions underpinning Davison's dissection of SBA. Such contrast enables readers to correlate Davison's arguments in 2007 with current realities. Most important, we can determine if Davison's findings remain valid in Hong Kong.

6. Limitations of Cheng, Andrews, and Yu's study

The implementation of SBA included a significant syllabus shift from the HKCEE to the Hong Kong Diploma of Secondary Education, which is entirely different in structure. The major difference was the duration of secondary school, which was extended from five to six years. Such a major reform contributed an extra SBA to the F.6 curriculum, increasing the total to three SBAs. Although the overall mark allocation of 15% for English papers was still the same, intrinsic factors trivializing the analysis remains ambiguous, and related subjects should be further explored.
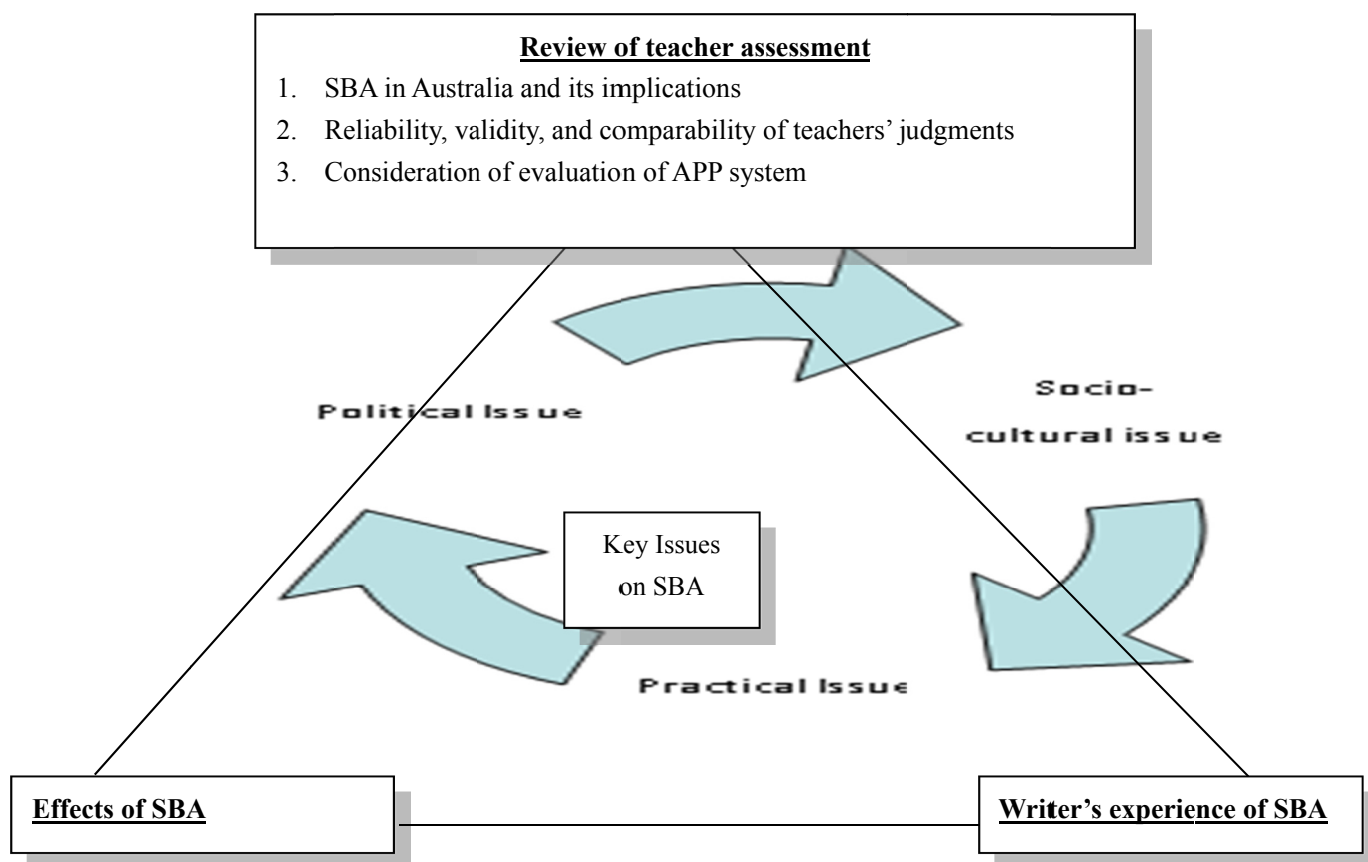
Review of teacher assessment
1. SBA in Australia and its implications
2. Reliability, validity, and comparability of teachers' judgments
3. Consideration of evaluation of APP system

Political Issue

Socio-cultural issue

Key Issues on SBA

Practical Issue

Effects of SBA

Writer's experience of SBA

Figure 1. Theoretical framework

7. Theoretical Framework of Present Analysis

7.1 First dimension: Standard-based teacher assessment in Australia and its implications

7.1.1 Implications of Australian system (university entry system)

The practices in New South Wales (NSW) reveal certain similarities between the Australian and Hong Kong systems. These similarities are comparable with Davison's interpretation of the imported SBA as influenced by Western countries in the new era of assessment (Davison, 2004). MacCann, Gardner, Reynolds, and Wild (2009) suggested that changes had occurred, such as the transformation of society from a labor-intensive structure to a knowledge- and service-based one, awareness of catering to social identity over the traditional sense of authority, structural changes in the nature of work, and the prevalence of informational technology. All these changes are also similar in Hong Kong, catalyzing the birth of educational reform, such as the introduction of SBA. However, apart from the influence of globalization, Davison only touched on how SBA was catalyzed by the above factors.

7.1.2 Assessment for learning

Both Davison (2007) and Stanley, MacCann, Gardner, Reynolds, and Wild (2009) have similar views that assessment innovation is migrating from formal, one-off summative assessment to a range of formative assessment tasks carried out and justified by teachers. Such tasks should be built on the authentic use of knowledge and should diagnose the weaknesses and strengths of assessment takers. Accordingly, a diversity of skills should be incorporated in integrated tasks. However, Davison (2007) indicated that students were solely asked to produce one individual presentation and participate in group discussion, which may be another one-off summative assessment that can be trivialized by the one-sidedness of teachers. Such a dilemma was echoed owing to the limited sources used for scoring. In a global level, as in Victoria and NSW, a wide range of sources were accounted for in the marking criteria (Stanley, MacCann, Gardner, Reynolds, and Wild, 2009). These sources included peer assessment forms, working portfolios, log books, teachers' observations, worksheets, and varied assignments. The acceptance of multiple sources readily enhances the reliability of tests, an effect that should have been considered by Davison.

Davison's paper also inadequately explained how teachers can effectively and substantially utilize their feedback to enhance their students' competence. An adequate explanation was intertwined with Davison's rationale and was thus expected. This opinion was supported by Annie, who proposed polarized ideas against this ideology of assessment for learning because several students do not value formative informal feedback, only feedback from summative assessments. Therefore, Davison should complement how teachers manage to provide students one-to-one, focused, and organized feedback with concrete suggestions and support for its improvement. More important, such feedback and comments should be precise, prompt, and constructive and are best provided in written form. In other words, the four aspects of comments, namely, quantity, quality, specificity, and clarity, should be interconnected in such a way that pronunciation, communication strategies, range of vocabulary, and sentence structures are clear and how to deliver ideas in a structural manner is sophisticated. In

addition, a supportive and welcoming learning environment advocated by teachers is a must to foster students' autonomy because formative feedback facilitates student progress and does not punish mistakes and ambiguity. Another crucial factor is the open-mindedness of teachers in terms of giving students differentiated feedback because of students' different perspectives of SBA learning outcomes (Annie, 2011). In addition, teachers should be alert with regard to giving feasible suggestions that can be incorporated into students' learning. Feedback concerning SBA performance considerably influences students' learning and potential improvement in SBA; therefore, such feedback should have been broadly discussed in Davison's paper.

7.1.3 Standard-based curriculum and assessment

Another crucial matter that Davison neglected was the linkage between what is expected and what is assessed from students. Davison (2007) argued that teachers should be given flexibility and space to customize learning tasks in addressing the various needs of students. This argument is consistent with the underpinnings of global practices, such as the practices in Victoria, NSW, and Queensland. However, teachers felt frustrated with the ambivalence caused by unconstrained flexibility. Neither did the teachers know what should be attained by students nor what corresponding activities should be included, leading to the decline of content validity. Davison could have suggested concrete and solid guidelines, standard handbooks, guidance, files, materials, and even learning kits to enhance achievability and comparability during the implementation (Kunnan, 1999; Cheung, 2001; Majid, 2011). Explicit cases can have certain standardization of learning tasks in the same school or district. Following this suggestion, Wong (2009) pointed out that the innovation of SBA is triggered by the social transformation in Hong Kong and the global trend in assessment reform. The momentum of this innovation is influenced by several crucial factors, such as the supply of resources, teacher training programs, the shift of the roles of teachers and students, and standardized tasks. To catalyze SBA implementation, task-based and small-class teaching can be adopted to counter intractability and teachers' dismay caused by the extra workload.

7.1.4 Expanded roles for teachers

Both Davison and Stanley, MacCann, Gardner, Reynolds, and Wild argued that expectations on teachers' duties and requirements are increasing. From the global perspective, Stanley, MacCann, Gardner, Reynolds, and Wild (2009) raised concerns on three dimensions of the expansion of teachers' roles. Regardless of the level (global or local), teachers are frontline assessors who evaluate their students' learning progress and achievements with great skill and transparency of imposed standards. In Davison's 2007 study, teachers were invited to construct the four domains of scoring criteria during the preliminary implementation. These domains were correlated with the global practices in 2009 (Hamp-lyons, 2009). With the very invaluable reassurance of teachers' judgments of their moderation, Davison could have advised on the feasibility of engaging teachers proactively in statistical and social consensus moderation. Another practical dimension of enhancing the reliability of teachers' moderation is extended inspection by external examiners, such as relevant experts in academics or other stakeholders (e.g., parents). Such enhancement would make Davison's assumption of the

expanded roles of teachers comprehensive.

Teachers' roles have been expanded, but teachers' beliefs that drive their motivation in taking up the expanded roles should not be ignored. Berry & Adamson (2012) pointed out that teachers' beliefs significantly influence the implementation of SBA. Teachers who support the value of SBA adjust and diversify the teaching content, whereas those who do not support SBA diminish the importance of formative assessment and customize few materials for SBA because the endeavor mainly focuses on the preparation of summative examinations. In other words, teachers' roles have undeniably been expanded, but the extent to which teachers' beliefs can be altered should have been prioritized by Davison over the newly launched assessment.

7.2 Second dimension: Reliability, validity, and comparability of teachers' judgments

7.2.1 Reliability (Consistency)

The notion that teachers interact with their students and judge their performance through a range of classroom activities every day implies that the reliability of their daily observations is higher than that of one-off summative assessment (Davison, 2007; Davison & Leung, 2009; Davison & Hamp-Lyons, 2009). However, crucial factors exist, such as theoretical, practical, and political issues and the purposes of the assessment. Stanley, MacCann, Gardner, Reynolds, and Wild (2009) found that the reliability of the Assessing Pupils' Progress (APP) assessment can be constructed through various measures collected over a period with consistent observation from teachers.

The triangulation for maintaining SBA reliability is constructed by the same teacher based on input under school-contextualized conditions throughout a long period longitudinally. However, Davison (2007) expressed his deepest concern by claiming that such evolving assessment must be supplemented with the mediation of reliability originating from psychometric criteria to strive for public acceptance. Davison's consideration of mixing the reliability of SBA and psychometric assessment is innovative and modest in the new assessment era. However, this idea underestimates the complexity of such synthesis. Moreover, as attempted in Queensland and other systems, diversified ways may be adopted to enhance reliability. These strategies include marking work samples and portfolios, augmenting assessment quantity, and establishing dual assessors (Stanley, MacCann, Gardner, Reynolds & Wild, 2009). These attempts are highly consistent and thus sustain high reliability. Therefore, these strategies may effectively compensate for the inadequacy of reliability in SBA advocated by Davison.

7.2.2 Validity (Accuracy and Adjustment)

In accordance with Stanley, MacCann, Gardner, Reynolds, and Wild's study, validity in assessment can be categorically construed as follows: (1) content validity, which refers to how well the assessment task is correlated to learning content fixed in the curriculum; (2) criterion-related validity, which is measured by the comparison of scores obtained in the assessment task within certain criteria; (3) concurrent validity, which indicates how an assessment task evaluates a student's performance; and (4) construct validity, which refers to

the qualities that a test uses to determine students' performance.

Regarding constructs and measured variables, Davison (2007) also proposed a set of four assessment criteria consisting of six descriptors in each domain customized by teachers and students during the piloting procedure. With each domain having the same mark allocation, the four domains are combined to form holistic assessment constructs and measured variables. Davison must have realized that SBA fulfills criterion-related and construct validity for the domains and thus clearly shows what criteria students should meet and how marks are allocated in each domain. In addition, SBA also exhibits concurrent validity because it assesses students at various stages from F.4 to F.5.

However, because of the absence of concrete and solid guidelines that show what should be taught or what learning exercises should be accomplished to attain promising results, content validity is reduced. In this regard, Davison should have further suggested what learning and relevant learning tasks can be incorporated so that what is taught is clearly assessed. External moderation may also assist in maintaining validity prior to the final standardization. As a type of APP, SBA is embedded in the authentic use of English. Therefore, how to measure and attain this area is the heart of assessment validity. On the flip side of validity, Hamp-Lyons (2011) argued that the assessors of SBA are under scrutiny with regard to how they can fairly assess their students and how their practice is related to the assessment. Such transformation imposes an unprecedented stringent requirement on teachers. In this regard, teachers should not remain at a loss but should adapt themselves to the assessment innovation. Teachers should regard this shift as a new acquisition process.

Another practical solution for Davison's study concerning validity can be rendered with the assistance of technology. Brown (2011) argued that SBA implementation is fully intractable, particularly to the extent of upholding its validity because it neglects the justifications on the portfolios, peers, and self-assessment and other performance. He further claimed that practice can lack validity because not all teachers are experts in the assessments. To compensate for such insufficiency, Brown (2011) suggested that computer assistance is necessary for mark analysis so that teachers can focus on facilitating an authentic learning environment in school.

7.2.3 Comparability

Apart from reliability and validity, Stanley, MacCann, Gardner, Reynolds, and Wild (2009) highlighted the importance of comparability, which refers to how marks, grades, or achievement indicators are evaluated on a common scale. Although Davison did not explicitly refer to or define comparability in his study, he pointed out that assessors comment on students' performance with the assessment forms and video-taping to finalize marks. F.4 and F.5 teachers then standardize their finalized marks in a coordination meeting, and a representative from each school then reviews the performance of other schools as a reference. All relevant materials, including assessment records, score sheets, and video-taping CDs, are then submitted to the HKEAA for final standardization and statistical moderation to assure stakeholders of the comparability of scores. Davison informed readers to some extent that comparability is introduced throughout the implementation. However, how comparability can be attained is not specified. For instance, how teachers guarantee that their marks are given

within the scale, how school-district coordination meetings assist panel heads or SBA coordinators in ensuring comparability in the school, and how the HKEAA standardizes and finalizes SBA scores are not explained. Further, Davison should have backed up his conceptualization of SBA comparability by reinforcing the monitoring of teachers during moderation meetings so that their standardization becomes obligatory. Teachers mediate any assessment innovation (Kunnan, 1999). For the HKEAA, the process of how the examination authority moderates students' results can also be disclosed as another reference concerning comparability.

7.2.4 Practical issues on managing assessment

In the global level, tensions among teachers are unavoidable during the implementation of the assessment process (Kunnan, 1999; Cheung, 2001). These tensions relate to the amount of work teachers should judge, the practicality of providing judgments in their pedagogy, and the proper and safe storage of assessment records. Teachers might be dismayed to conduct too many observations and write correspondent reports.

In addition to such tensions, Davison (2007) claimed that most secondary school teachers report dismay and are reluctant to conduct SBA in school. Teachers are deeply concerned with their inadequate skills and knowledge related to the implementation of SBA and are worried that they might be overwhelmed by the extra workload from SBA, such as the customized learning activities, score collection, and routines in running the actual assessment. Another disadvantage that Davison pointed out was the limited time and resources given to teachers for their adoption and adaptation, leaving them with only partial understanding of SBA philosophy and values. Worse, EDB had neither alleviated their apprehension caused by the ambivalence nor given encouragement and modeling to the struggling teachers (Wong, 2009; Berry & Adamson, 2012).

Therefore, the global practices are hardly comparable with those in the Hong Kong setting in terms of practical issues because of the different nature of the assessment of portfolios and presentations. One common area is teachers' concerns about the overwhelming workload in judging students' performance. Thus, Davison's study can be supplemented with possible solutions that both reduce teachers' workload and maximize their capacity and versatility for implementation. These solutions might be the best remedy for the neglect of giving appropriate advice on these issues.

For instance, the Education Bureau can establish an electronic database that stores certain videotapes of students' levels and is refreshed regularly to enable referencing and moderation by teachers (Brown, 2004). Moreover, teachers should receive sufficient training on campus and various workshops organized by experts in the assessment field. Such programs must be based on practicality and real engagement with SBA materials in an authentic context. Ideally, such training can be initiated in the classroom with some trials to evaluate students' performance. Another auxiliary solution can be established with an apprenticeship scheme, which invites experienced and senior teachers to impart their valuable experiences to new teachers to sustain good teaching practices and to provide advice to any staff newly exposed to the assessment innovation. Apart from apprenticeship, middle managers of schools should

arrange a suitable assessment schedule to cater to the needs of teachers, who are spared from extra workload particularly during students' presentations. Administration staff, such as the IT crew, should always be equipped and willing to provide any technical support, such as audio/video equipment, as well as effective methods to store video clips. In a wider scope, the Education Bureau can recruit additional English teachers to reduce the burden derived from the assessments. The ultimate success of SBA lies in the assessors. Thus, by enhancing their sense of ownership, professionalism, and motivation, teachers incline toward better commitment and involvement in SBA (Kunnan, 1999).

These suggestions are backed up by Lo and 羅耀珍 (2006), who suggested that small-class teaching serves as a crucial factor for SBA implementation because assessors (teachers) can be given more space and time to focus on individual students, especially those coping with learning difficulties. In an in-depth investigation, the government should tailor SBA to cater to the needs of SEN students.

7.2.5 Cheating and authenticity of student work

Stanley, MacCann, Gardner, Reynolds, and Wild (2009) suggested that assessment innovation possibly provokes concern on any attempts on cheating and the scrutiny on the authenticity of student work. However, one of the most rigorous debates is placed on how teachers help students enhance their academic results by preparatory tasks and exposure to the actual assessments. The inconsistency or inappropriateness of teaching practices might disadvantage students in their performance (Yung & Yung, 2001). However, Davison only touched on the inconsistency or inappropriateness of tasks. Davison could further suggest the documentation and legitimization of effective practices prior to implementation to maintain the validity of teachers' compliance. Another possible enhancement is the internal and external scrutiny of teaching practices and materials. Internally, head teachers and panel heads should periodically check teachers' worksheets, exercises, and teaching plans concerning SBA. External inspection can be exercised by professional communities, such as academics, teacher unions, and even relevant sectors in Hong Kong. In case of appeals from parents and students, schools and assessors should be open and prompt to explain any discrepancies to resolve grievances. All these approaches can complement Davison's study to guarantee that the assurance of the authenticity of teaching practices is comprehensive and linear.

7.2.6 Plagiarism

Stanley, MacCann, Gardner, Reynolds, and Wild (2009) raised the concern on whether submitted assignments are plagiarized through downloading, paraphrasing, and copying of information directly from the Internet without proper acknowledgment and referencing. In addition to plagiarism, the public is becoming skeptical of who actually produces the assignment or to what extent students are offered intervention from outsourced providers to complete their work (Cheung, Hatte, Bucat & Douglas, 1996).

Davison (2007) indicated that the ideology behind SBA is to make sure that students do not recite phrases and sentences from a prepared script but that they produce the best of their own

work. Therefore, he claimed that assessors should prohibit students from plagiarism by memorizing pre-written lines. Technically, Davison did not respond to the issue of anti-plagiarism because he offered no concrete solution to prevent students from plagiarism. Undoubtedly, because the examination venue is videotaped, recitation of lines from a completely furnished script is ineligible. However, the originality of the work is always somehow doubtful. Students can always copy and paste information from Wikipedia, blogs, and critiques for individual presentations and group discussions without the awareness of teachers. Without proper software or technology, teacher inspections are arguably only in vain. Another problem is how teachers identify that the work is accomplished solely by students. Regardless of the fact that students mentored by a teacher might be equally assessed in terms of quantity and quality, a counterintuitive argument reveals that students with better resources benefit from various outsourced providers. Davison should strengthen his ground on avoiding plagiarism with the following suggestions: (1) using anti-plagiarism software, such as Turnitin, which scrutinizes any plagiarism on the web; (2) enacting strict laws to penalize assessor takers who commit plagiarism and assessment users who violate ethics; and (3) advocating moral education to students, parents, and teachers on how plagiarism adversely affects individuals and society.

7.3 Third dimension: Consideration of evaluation of APP system

7.3.1 What the evaluation could comprise

Stanley, MacCann, Gardner, Reynolds, and Wild (2009) pointed out that evaluation of the assessment is the key to its sustainability and extensibility and that its success heavily depends on how classroom routines vary before and after the assessments. Some parameters might include the evaluation of the efficacy of scoring criteria and domains, the storage of data, the accountability of assessment tasks, the achievability of global standards, and most significant, the workloads of teachers. Thus, a comprehensive yet simple and open evaluation with interviews and questionnaires is inevitably required.

Davison partly explained how SBA was initiated and molded with the interviews and questionnaires collected by the Faculty of Education of the University of Hong Kong in alignment with the HKEAA. Test users and takers were assumed to actively engage in evaluation prior to the actual assessment.

However, Davison did not describe the post-implementation aspect of the evaluation, which was vital to the completeness of his study. Without effective evaluation strategies, SBA is just a one-off summative assessment because the whole operation system is a closed black box when it is not revitalized with ideas from various stakeholders (Cheung, Hatte, Bucat & Douglas, 1996). To fill the gap in Davison's evaluation, we suggest various questionnaires to discern issues related to the assessment and, when necessary, follow-up questionnaires and in-depth interviews to further address dilemmas, grievances, and dissatisfaction. The whole process can be divided into two phases. Likert-scale questionnaires are collected from teachers, students, and parents so that substantial data can be computed and processed in a prompt and organized manner. The bottom of the sheet should have some space for comments because some issues might be excluded by the fixed and structured items in the

questionnaires. With both qualitative and quantitative bases covered, the feedback can reflect the performance of the administration and school so any possible diagnostic can be counter-established efficiently (Stanley, MacCann, Gardner, Reynolds, and Wild, 2009) . From the obtained responses, customized questionnaires might be deployed to further the researchers' understanding of both positive and negative responses to investigate the complexity of varied issues. To determine stakeholders' perspectives about causes and effects, structured interviews are required to hear from students, teachers, and even parents to enhance the breadth of the whole evaluation, with which the fine-tuning of the assessment content and context becomes accurate. With this prerequisite, differentiated questionnaires should be designed and oriented to a range of stakeholders and administrators, including teachers, senior teachers, panel heads, IT technicians, parents, students, head teachers, and administrative staff, whose opinions can be refracted from the prism of this comprehensive yet strong evaluation mechanism.

8. Evaluation of Davison's findings on SBA implementation

Cheng, Andrews, and Yu's study is used to serve as the prism refracting Davison's comprehension of SBA with regard to the preliminary stage of SBA and to supplement the insufficiency of his study.

8.1 Effect of SBA on students

Davison (2007) suggested that in accordance with the interview and questionnaires, students generally hold positive attitudes toward the underlying philosophy and values of SBA because they realize that they can enhance their oral competence in a naturalistic and contextualized context by means of a range of simulated assessment activities. Cheng, Andrews, and Yu (2011) indicated that students with various English proficiencies construe SBA in different ways. Highly competent students focus on external examination results but are only slightly concerned about any anticipated routines, activities, and challenges of SBA (Annie, 2011). In other words, the ideology of transforming the assessment of learning to assessment for learning was in vain because of proficient students' ignorance of perceived challenges, contradicting Davison's assumption that it is teachers who mediate SBA by forfeiting assessment practices for students and consider SBA as a one-off assessment. Conversely, Davison's assumption is valid only with low-proficiency students. One purpose of SBA is to scaffold and boost low- and intermediate-level students' confidence and competence for their achievement. As a result, Davison's study can be improved by suggesting that the HKEAA should address the issues of students' differences in English skills and refrain from being unilateral to make the assessment innovation universally beneficial.

Further, Cheng, Andrews, and Yu (2011) found that the nearer the day of the assessment, the more frequently activities are carried out, implying that the school should always schedule learning activities according to the timeline of the assessment to maximize the efficacy of student learning. This finding can be extended to future research as suggested by Davison.

8.2 Effect of SBA on parents

Davison focused on surveying teachers' opinions and foundational knowledge about SBA and thus neglected the parents, who are equally crucial stakeholders as teachers and students. Cheng, Andrews, and Yu (2011) found five interlinked factors influencing SBA implementation: parents' socioeconomic background, the time parents spend with their children, parents' knowledge of SBA, their perceptions about SBA, and their opportunities to know about SBA. These factors are substantive and substantial to their support for their children in relation to SBA. Cheng, Andrews, and Yu found that among these factors, parents' knowledge about SBA is the most influential on parental support. With this finding, Davison could have reminded schools of their obligations to advise parents on how SBA is implemented in school so that they know how to provide their children with auxiliary support when appropriate. With the triangulation of responses from parents, students, and teachers, Davison's investigation of the effects of SBA will become sounder.

9. Writer's Experience of SBA

As the English panel head in a band-three secondary school serving many dyslexic students for six years, the writer has encountered quite an amount of perplexity, ambiguity, grievance, and desperation in the implementation of SBA. First, no clear instructions are given in SBA to cater to the needs of SEN students, although Davison (2007) and Stanley, MacCann, Gardner, Reynolds, and Wild (2009) have suggested the importance of maintaining fairness in the "playground level." In fact, most dyslexic students encounter great difficulty in decoding texts. These students can neither deliver individual presentations in a smooth flow nor successfully initiate group interaction cohesively even when teachers provide them with remedies and interventions regularly. SBA accounts for 15% of the overall mark allocation, and students do exert efforts in preparing their presentations, but because of their inherent inadequacy in decoding text, they can only copy information from the web or plagiarize from others although they know that they would be disqualified once found copying. Some students choose to take risks, and some others give up. Every year, two or three candidates drop out. Second, teachers are frustrated by the different groupings prior to group discussion because unmotivated students rarely contribute to the preparation, increasing complaints from other group members. Teachers are also impeded by the insufficiency of relevant reference when preparing the assessment practice (Davison & Leung, 2009). Teachers comply with the timeline fixed by the HKEAA in accordance with the handbook, which proposes only few relevant simulation activities. Occasionally, workshops and sharing sessions are held to improve teachers' skills and testing strategies. Rarely do organizers or the school principal consider the overwhelming workload of teachers. Thus, interested teachers either decline the training offer or swap lessons to attend the training, which in fact substantially adds to the teachers' workload, even though teachers appreciate the underlying philosophy and values of SBA. In addition to students' disabilities and teachers' overwhelming workload, low-income parents cannot assist their children in their studies or in SBA because of limited time and restricted knowledge. Only a few parents are aware of the setting of SBA (Cheng, Andrews,

and Yu, 2011). In sum, band-three students, particularly those with SEN, are disadvantaged in SBA, contradicting Davison, who assumed SBA to be most beneficial to low- and intermediate-level students.

This counterargument is supported by Annie's (2011) case analysis of how students and teachers perceive SBA; such perceptions are controversial. High-banding-school students claimed that SBA enhances their speaking and presentation skills with regular practice and videotaping. However, most teachers, particularly those in band-three schools, are overwhelmed by the lack of motivation and ability of their students, who might just copy from Internet sources. Moreover, some teachers reported that without enough resources, extra teaching loads from free periods are imposed on them to practice SBA, and panel heads are burned out with additional internal workshops and briefing to teach new teachers the concepts of SBA. Fundamentally, teachers performing not what are expected but something irrelevant and complicated.

10. Conclusion

Parents are uninformed and ignorant of SBA implementation even though it is a high-stakes test. These findings contradict Davison's assumption that SBA is widely accepted and appreciated by various stakeholders. The value of SBA is inarguable. Berry and Adamson (2012) noted that informal alternative assessments have become inevitable because the teacher-centered approach is invalid owing to the structural change in the socio-economy and the emphasis on exploring individual potential against the dominant value embedded in tradition. Nonetheless, the balance should be maintained to ensure that the assessment practice is both widely accepted and sustainable. Such balance can be achieved by empowering both teachers and students to utilize self-assessment and peer assessments to reduce the subjectivity of one-sidedness. Generally, Davison can reassure stakeholders of two practical ways to ensure reliability, validity, and comparability. First, the Education Bureau should offer feedback to schools so that teachers and panel heads can proactively address their strengths and weaknesses in their teaching practices through self-evaluation. Second, assessment preparatory tasks and practices should be standardized. Although such standardization might impede teachers' flexibility in tailoring learning materials, it is advantageous to both assessment takers and assessment users.

References

ANNIE, T. S. Y. (2012). Exploring Students' Perception of and Reaction to Feedback in School–based Assessment.

Berry, R., & Adamson, B. (2012). Assessment Reform in Hong Kong Schools. *Assessment*.

Brown*, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318. http://dx.doi.org/10.1080/0969594042000304609

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing, 28*(2), 221-249. http://dx.doi.org/10.1177/0265532210384253

Cheung, D. (2001). School-based assessment in public examinations: Identifying the concerns of teachers. *Education Journal-Hong Kong-Chinese University Of Hong Kong, 29*(2), 105-123.

Cheung, D., Hattie, J., Bucat, R., & Douglas, G. (1996). Measuring the degree of implementation of school-based assessment schemes for practical science. *Research in Science Education, 26*(4), 375-389. http://dx.doi.org/10.1007/BF02357450

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305-334. http://dx.doi.org/10.1191/0265532204lt286oa

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37-68. http://dx.doi.org/10.1080/15434300701348359

Davison, C., & Hamp-Lyons, L. (2009). The Hong Kong Certificate of Education: school-based assessment reform in Hong Kong English language education.

Davison, C., & Leung, C. (2009). Current Issues in English Language Teacher‐Based Assessment. *Tesol Quarterly, 43*(3), 393-415. http://dx.doi.org/10.1002/j.1545-7249.2009.tb00242.x

HAMP‐LYONS, L. (2009). Principles for Large‐Scale Classroom‐Based Teacher Assessment of English Learners' Language: An Initial Framework From School‐Based Assessment in Hong Kong. *Tesol Quarterly, 43*(3), 524-530. http://dx.doi.org/10.1002/j.1545-7249.2009.tb00249.x

Kunnan, A. J. (1999). Recent developments in language testing. *Annual Review of Applied Linguistics, 19*, 235-253. http://dx.doi.org/10.1017/S0267190599190123

LO, Y. C., & 羅耀珍. (2006). Practice and challenges of school-based formative assessment.

Majid, F. A. (2011). School-Based Assessment in Malaysian Schools: The Concerns of the English Teachers. *Online Submission*.

Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). *Review of teacher assessment: Evidence of what works best and issues for development*. London: Qualifications and Curriculum Authority.

Wong, M. L.-Y. (2009). Perspectives on the English Language Education of Hong Kong's New Senior Secondary (NSS) Curriculum. *Professional Teaching Articles, 35*, 1-27.

Yung, B. H. W., & Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education, 23*(10), 985-1005. http://dx.doi.org/10.1080/09500690010017129

Copyright Disclaimer