

Washback Effect of School-based English Language Assessment: A Case-Study on Students' Perceptions

**Alla Baksh, M. A.^{1*}, Mohd Sallehuddin, A. A.², Tayeb, Y. A.² and
Norhaslinda, H.³**

¹*School of Languages, Literacies and Translation, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia*

²*School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

³*Akademi of Language Studies, Universiti Teknologi MARA, Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang, Malaysia*

ABSTRACT

This paper reports on the preliminary findings of an on-going study on the washback effect of the newly introduced school-based assessment (SBA) at the lower-secondary level in Malaysia. This study specifically investigates how the school-based assessment has affected the perceptions of students in relation to learning English as a second language. In addition, the study attempts to explore the students' responses to a call for change from a purely testing culture into a learning culture at the beginning of its implementation. The objectives of the study are therefore twofold: *to gauge the washback effect on students' overall perceptions of SBA and external examinations and challenges of implementing School-based Assessment (SBA)*. Drawing on the data collected by means of questionnaires, it was found that the sampled students were equally pessimistic about external examinations and SBA. In addition, some barriers in implementing SBA as perceived by the students in the given context are reported. It is hoped that the findings of this small-scale study which was carried out after two years into the implementation of SBA, would contribute to a better understanding of the complex phenomenon of "washback" in relation to SBA in Malaysia.

Keywords: Washback effect, school-based assessment, students' perceptions, English language, lower-secondary level, low-stakes, Malaysia

ARTICLE INFO

Article history:

Received: 14 May 2015

Accepted: 18 May 2016

E-mail addresses:

allabaksh@usm.my (Alla Baksh, M. A.),

salleh@ukm.edu.my (Mohd Sallehuddin, A. A.),

yahyaameen73@yahoo.com (Tayeb, Y. A.),

haslinda.hassan@ppinang.uitm.edu.my (Norhaslinda, H.)

* Corresponding author

INTRODUCTION

Malaysia as one of the British colonies has adopted many of its administrative systems which includes its education system (Saw, 2010). Students in Malaysia undergo 11 years of compulsory schooling which requires them to sit for three major public examinations. The first public examination i.e., Primary School Assessment or the UPSR (Ujian Penilaian Sekolah Rendah) is carried out in the sixth year (end) of the primary level. The lower secondary assessment i.e., the focus of the study, was initially known as *PMR (Peperiksaan Menengah Rendah) before it was named as Pentaksiran Tingkatan 3 (PT3) or Form 3 assessment in the year of 2014, was the next public examination conducted at the end of lower-secondary level (year 9) till 2013 and the third public examination is the Malaysian Certificate of Education or the SPM (Sijil Pelajaran Malaysia) which is carried out in the fifth year of secondary level (year 11).

The assessments at the national schools till the mid-nineties were centralised as they were entirely handled by two central bodies namely the Malaysian Examinations Syndicate (*MES*) and the Malaysian Examinations Council (*MEC*). The aforementioned standardised public examinations at the primary (*UPSR*), lower-secondary (*PMR/PT3**) and upper-secondary (*SPM*) levels in schools fall within the jurisdiction of the Malaysian Examinations Syndicate (*MES*) in ensuring their validity and objectivity (Chan et al., 2009; Baidzawi & Abu, 2013). The Malaysian Examinations

Council (*MEC*) on the other hand, exercises similar jurisdiction over the Malaysian Higher School Certificate or the STPM (*Sijil Tinggi Persekolahan Malaysia*) and MUET (*Malaysian University English Test*) which are sat by sixth-form or the pre-university (*year 12 and 13*) students who are in their final stage of school education before gaining entry into higher learning institutions either at local or international institutions acknowledged by the Malaysian government (Saw, 2010).

Considering the overarching examination oriented-culture which had been in practice for over 30 years, the Malaysian government embarked on a major assessment reform at both primary and lower-secondary levels of education in 2011 and 2012 respectively. An entirely school-based assessment shifting the paradigm of teaching duties of teachers from 'teaching only' into a 'teaching and assessing their own students' at both levels were introduced. Hence, there are various implications for the students from such a transition in the assessment system whereof they may need to make necessary adjustments in 'what' and 'how' they prepare for the newly-introduced assessment system.

School-based assessment was introduced at both primary and secondary levels. However, considering the years and levels they were introduced, the researchers had to investigate the lower-secondary level assessment as the first batch of students at this level were about to experience the new assessment system (*PT3*) compared with their counterparts at the primary level

who will experience a revised assessment system in 2016. Thus, the scope of this paper is on the English language assessment reform at the lower-secondary level only. Specifically, it attempts to gauge the consequences of the newly-introduced English language assessment system on the ultimate stakeholders (of any assessment) as claimed by Bailey (1996) - the students. These consequences of assessment affecting teaching and learning at the micro level (classroom) is referred to as 'washback' or 'backwash' by language testing scholars. Hence, this study looks into the washback effect of a school-based English assessment on the perceptions of a group of students at the lower-secondary level in one of the schools in Penang.

First, this paper briefly introduces the construct of school-based assessment and discusses the type of school-based assessment practised at the lower-secondary level in Malaysia. The language testing construct "washback" is then introduced and is linked with the newly-introduced school-based assessment in Malaysia. Next, methods and instruments employed for the study are discussed followed by significant results and discussion. Finally, conclusions and some recommendations are drawn along with the limitations of this study.

SCHOOL-BASED ASSESSMENT IN MALAYSIA

Standardised public examinations at different stages of education are prevalent in many education systems around the globe. However, a review of literature

on language assessment reveals that traditional assessment systems in which the policymakers (stakeholders at the macro level) judging the efficacy of language teaching and learning in classrooms by means of student achievement in one-off standardised summative tests do not always have a beneficial impact. In this regard, Alderson and Wall (1993) have clearly disputed the claim made by scholars like Morrow (1986) that 'a test is valid when it has good washback and it is invalid when it has negative washback'. In today's reality, many education systems have the examination questions constructed by policymakers who do not teach the subject and teachers who teach the subject, not being directly involved in constructing the exam questions. Notwithstanding, the stakeholders at both macro (policymakers) and micro (teachers and students) levels should make every effort to harmonise the relationship between the curriculum, teaching and learning activities and the assessments which may lead the stakeholders in achieving positive washback (Shohamy et al., 1996). Unfortunately, this does not happen as has been proven by many empirical studies carried out in various contexts to date. Hence, systems of this kind have mostly pressured the students in choosing between either to pass the test or improve their language proficiency (Wall, 1996; Qi, 2005).

Therefore, education specialists around the world in realising the shortcomings of such one-off standardised tests have gradually begun looking into the gaps

identified in the approaches of assessments employed in the 20th and addressing them in 21st century. “While the former pursued the evidence at the end of the learning process (summative), the international agenda for twenty first century is the recognition of using assessment for learning purposes (formative)” (Berry, 2011).

In line with this global shift in assessment practices, the Malaysian government introduced standards-referenced school-based assessment into its education system at the primary and lower secondary levels in 2011 and 2012 respectively. The rationales and stages of implementation were discussed in the Blueprint (2013).

Shohamy (1991) argued that the use of tests for power and control is a very common practice in countries which the educational systems are centralised: the curriculum is controlled by central agencies. Malaysia is one of these countries which controls its curriculum, teaching and learning, and assessment through MES and MEC. The government’s intention of implementing SBA is to promote real learning of the subject matters among the students instead of rote-learning and memorisation (MES, 2014). However, given the stakes attached to assessments at different levels along with the society’s (macro-level stakeholders) faith in teachers grading their own students without fear and favour, the Malaysian government had to choose the lower levels of education namely primary and lower-secondary levels in implementing an entirely school-based assessment in which the role of central

agencies is minimised but the teachers’ role as assessors is increased.

On the other hand, Malaysia has actively been participating in international assessments like PISA (*Program for International Student Assessment*) and TIMSS (*Trends in International Mathematics and Science Study*). A below-average performance by 15 year old Malaysian students in such international assessments was another factor that led the government to consider and embark on assessment reforms. It was discovered by means of recent PISA results that the Malaysian students were not able to deal with items which required higher-order thinking skills. Therefore, the Malaysian government has set a target of being ranked top-third in such international assessments by 2025 (MES, 2014).

Comparatively, implementing an entirely school-based assessment system with a minimal involvement of central agencies in a high-stakes test like SPM may not be accepted by society due to the impact the tests have on students’ future: scholarships and other perks are awarded based on the test results. However, in its effort to promote assessment for learning at this level, the government has implemented school-based oral assessment (SBOA) since 2002. Thus, considering the stakes/consequences of assessments on students’ lives along with the formative stage of learning of the students, SBA began with the primary and lower-secondary levels of education (low-stakes) (MES, 2014).

The assessment reform undertaken by Malaysia is a synergistic school-based assessment at the lower-secondary level (MES, 2014) in which four assessment components are contained within the new Form 3 assessment system.

They are:

- i. Form 3 (central) assessment
- ii. *School assessment*
- iii. Physical Activities, sports and co-curriculum, and;
- iv. Psychometric assessment

The first two assessments are categorised under the academic component whereas the other two are non-academic ones. At the end of the lower-secondary level, students now are provided with four different forms of results representing each component of the broad school-based assessment. The non-academic component of the school-based assessment is, however, beyond the scope of this paper.

The Form 3 (central) assessment is a summative paper-and-pencil test which involves an evaluation of all four language skills. Hence, it is assessment *of* learning as it comes at the end of the term. The MES provides the schools nationwide with sets of question papers to choose from (comparative standards). Teachers grade the exam scripts of their own students by strictly adhering to the guidelines provided by the ministry. It is, however, worth noting here that the previous assessment (*PMR*) focused mainly on the reading and writing skills only.

The school assessment on the other hand is a combination of formative and summative components. Three aspects are contained within the school assessment: assessment *for* learning, assessment *of* learning and assessment *as* learning. Researchers (Black et al., 2003, as cited in Yu, 2010) have opined that while raising students' achievement is the first priority of assessment *for* learning, it also involves teachers in multiple formal and informal assessment methods such as unit tests, quizzes, oral presentations, listening activities and homework to judge the quality of their students' learning against a set criteria or standards. In this regard, the MES has provided the teachers with a band scale of 1 to 6 in which 1 indicates weak and 6 indicates advanced learners. Students are required to achieve the highest bands possible over the year. Black et al. (2003, as cited in Yu, 2010) also highlighted that an assessment activity should, ideally, be able to provide feedback to both teachers and students which may assist them in assessing each other and also adapting their teaching and learning activities. The school assessment as stated in the official documents of the ministry (moe.gov.my), requires the teachers to identify their students' actual performance and their desired performance as required by the ministry (intended washback) and provide their students with necessary feedback with an aim to bridge the gap. Besides, students also have the opportunity to be assessed by their peers (peer-assessment) and themselves (self-assessment).

As mentioned earlier, Malaysia has implemented a one-off summative examination system at the lower-secondary level for about 30 years. Considering the recent shift in policies and practices in relation to the assessment at the lower-secondary level and the students who are one of the most affected key stakeholders, studies investigating their perceptions of such an assessment reform are deemed necessary.

A PANORAMIC VIEW OF WASHBACK

Learners' achievement in standardised examinations has been widely used as a tool to measure the performance of stakeholders at the micro level (*classrooms*), schools and educational systems (*administration*) for accountability purposes. Alderson and Wall's (1993, p.4) remark that 'tests are held to be powerful determiners of what happens in classroom' clearly supports this statement. The phenomenon of examinations influencing the teaching and learning activities is defined as washback in the area of language testing (Alderson & Wall, 1993). This phenomenon sparked an interest among scholars in both general and language education.

However, scholars, in exploring this phenomenon, have been divided in their definition. Though their definitions broadly deal with examinations influencing teaching and learning, there are some differences in relation to the scope of stakeholders affected by the examinations. Some significant definitions which have widely been reported

in the assessment literature are reported here and the definition identified for the study is then stated. Some scholars have argued that examinations may affect stakeholders at the micro level; classroom at the macro level; schools (administration), education systems and society at large. Bachman and Palmer (1996), Wall (1997), Andrews (2004) and McNamara (2010) refer to this phenomenon as 'test impact'.

Frederikson and Collins (1989) on the other hand introduced the term '*systemic validity*' which means effects of instructional changes brought about by the introduction of tests into the educational systems. Messick's (1996) *consequential validity* revolves around concepts ranging from the uses of tests, the potential misuse, abuse, and unintended usage of tests, the impacts of testing on test takers, teachers and the decision makers. Popham (1987) introduced the term *measurement-driven instruction*, Shohamy et al. (1996) defined *curriculum alignment* as altering the curriculum in line with the examination results and Morrow (1986) introduced *washback validity* which refers to the value of the relationship between the test and any associated teaching.

This family of terms have all one thing in common: curriculum, teaching and learning are controlled by means of either introducing new or altering existing examinations within the education systems. Washback or backwash is broadly defined as the effects of tests on teaching and also on learning (Cheng & Curtis, 2004). According to Alderson and Wall (1993), the term washback is widely used in British Applied

Linguistics but backwash is prevalent in educational literature. After reviewing the existing literature on the available definitions and taking into account the context in which a different assessment system has just been introduced, the present researchers adopted the term 'washback' propagated by Alderson and Wall (1993) at the micro level, classroom, to gauge the kind of washback on a group of learners' perceptions at the beginning of implementing a new assessment system in the current context. Therefore, this study uses the term washback to be used throughout as it deals with language education.

THEORETICAL DISCUSSION

Scholars (Morrow, 1986; Frederikson & Collins, 1989; Khaniya, 1990) from both general and language education have widely asserted the existence of washback without any empirical evidence. Washback in language testing domain came into prominence in early 1990s when Alderson & Wall (1993) disputed the assertions made by other scholars over the years that a good test will produce beneficial teaching and learning (positive washback) and vice versa. They argued that a test alone may not be the reason for the kind of teaching and learning observed in a language classroom but there might be other factors within classrooms, schools, educational systems and society at large at work which hinder washback from happening. They subsequently proposed 15 washback hypotheses in their seminal paper "Does washback exist?" which deal with 'what' and 'who' are affected

by tests. The 'whats' according to them are teaching - rate, sequence, degree and depth of teaching, and, learning - rate, sequence, degree and depth of learning and the 'whos' are teachers and learners. Hughes (1993) in his attempt to enhance the understanding of backwash (as he referred to it), broke the consequences down into three broad categories: participants, processes and product. Bailey (1996), combining both Alderson and Wall's (1993) and Hughes' (1993) insights, presented the ideas with an addition of 'researchers' into the participants' category in the form of a diagram (see figure 1)

Bailey (1996) propagated Hughes' (1993) message in her diagram that learning is the ultimate goal (product) of any assessment introduced into education systems. Hence, beginning anywhere in the diagram would eventually lead us to the 'learning' construct. The straight arrows in her diagram refer to the intended washback as required by policymakers and other beneficial consequences (research results). On the other hand, the dotted arrows refer to the washback effect observed within specific territories (teachers, learners, etc.) in relation to their perceptions, attitudes, teaching methods, learning strategies and so on. Finally, the product of the assessment system as claimed by Hughes and Bailey is the learning of the language.

The present study adopted Bailey's (1996) model to look into the washback effect of a school-based English language assessment on a group of learners' perceptions in one of the schools in a

northern state of Malaysia. In reference to Bailey's model, the participants involved in this study are the students and the processes of the students are as stated below:

- I. Students' overall perceptions of school-based assessment (SBA) and external Examinations.
- II. Challenges of Implementing School-based Assessment (SBA).

Washback on Students' Perceptions

It was argued in empirical washback studies that the perceptions of the key stakeholders (teachers and students) were among the significant factors which greatly influenced the teaching and learning activities in classrooms. Yu (2010) in her study on

the washback effect of the school-based oral assessment reported that the students had little knowledge of school-based assessment and they did not perceive any benefits that SBA claims to bring to learning (teacher feedback and peer-assessment). Considering the standards-referenced school-based assessment recently introduced by the Malaysian government at the lower-secondary level to promote learning, an investigation into the perceptions and attitudes of students in relation to the new assessment system is therefore necessary. The paucity of students' perspectives on the washback effect as reported in the literature internationally (Hamp-Lyons, 2000; Stoneman, 2006; Shih, 2009) and the need to know what is intended by the ministry

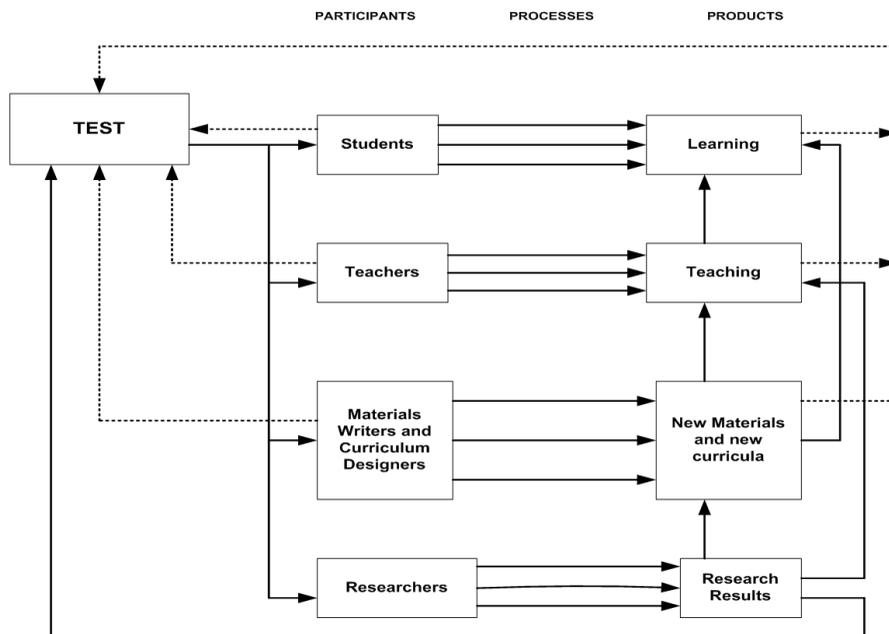


Figure 1. A basic model of washback (Bailey, 1996, p.264).

and what is happening in reality within classrooms warrants an investigation into the washback effect of the newly introduced school-based assessment on learners at the beginning of its implementation.

As the previous studies on school-based assessment in Malaysia have been centred on concerns among teachers with regards to its implementation (Faizah, 2011; Baidzawi & Abu, 2013; Nair et al., 2014), it is deemed significant and timely to carry out a study on the washback effect of school-based assessment on students' perceptions as they are one of the direct stakeholders of any assessment reforms.

Shih (2009) in her study on 'how tests change teaching?' highlighted that research on washback to date has been centred on washback of tests on teaching or they investigated the impact of teachers' educational backgrounds or beliefs on their teaching. However, scant attention has been paid to the role played by student factors in affecting teaching within the washback mechanism, and the researchers believe this is particularly so in Malaysia. According to Shih (2009), some of the potential areas of washback in relation to student factors are like 'how do students' feedback affect teaching?' and 'how do students' learning motivation influence teaching?'

METHOD

Participants

The participants of this study were from the lower-secondary level i.e., PT3 of a co-education national secondary school in Seberang Perai Tengah, Penang. The state

of Penang and the district of Seberang Perai Tengah were chosen for convenience purposes. The sampled school had requested anonymity. The researchers were able to sample a balanced number of both genders, 17 males and 17 females ($n=34$) from this school. Schools in Malaysia are ranked by bands: band 1 being the highest and band 6 the lowest. As this study was conducted at the very beginning of the implementation of SBA in Malaysia, the researchers employed purposive sampling by identifying a middle-band (band 4) school.

Choosing a school of middle banding to some extent minimises students' ability as a factor in making reference of the school's experiences to other contexts, as compared to the other two scenarios of either choosing a high-banding or a low-banding school. Experiences of a middle-banding school allow a wider range of readers to recognise similarities of issues in their own context (Yu, 2010, p.79).

Instruments

This study looks into the washback effect of a newly introduced language assessment. Therefore, the researchers had to rely on official documents (*booklets, press release, etc.*) issued by the ministry to gauge the positive/intended washback. After reviewing all the available documents, the researchers then triangulated the students' responses by means of their self-reported questionnaires. Therefore, this study

includes document analysis and survey as its research instruments.

A validated questionnaire by Yu (2010), who conducted a mixed-methods case study on the washback effects of school-based performance assessment in a Hong Kong secondary school, was adapted by this study. As the context for the present study is Malaysia, necessary amendments were made to suit the context. After thoroughly analysing the instrument, some items deemed not relevant were removed and where necessary, some new items were added. In addition, as the researchers' aims were to gauge whether the new assessment system has positively or negatively affected the students' language learning, the researchers in their attempt to avoid fence sitters had to transform the questionnaire originally designed on a 6-point Likert scale into a 4-point Likert scale. The respondents were required to respond on a 4-point Likert scale ranging from 1 which indicates *strongly disagree* to a score of 4 which indicates *strongly agree* for section II, IV and V whereas they were required to respond to a 4-point Likert scale ranging from a scale of 1 which indicates *never* to a scale of 4 which indicates *always/all the time* for section III. Only the results of section I and V are discussed in this paper.

The adapted questionnaire was revalidated by two local experts in the area of language testing and a reliability test was run for each item and for the entire instrument. An internal consistency test of the questionnaire revealed that its cronbach alpha value was at .88.

Data Collection and Analysis

The questionnaires were distributed by one of the teachers in the sampled school. They were completed by the students under the teacher's supervision with a return rate of 94%. Some questionnaires were not returned due to the absence of students on the day when the instrument was administered and some students failed to return their questionnaires. The Software Package for Social Sciences (SPSS, V21) was used to analyse this study data. An independent-samples t-test and the measure of central tendency were carried out to see the differences among the sampled respondents.

RESULTS

This study only reports the findings from sections I (*Students' overall perceptions of school-based assessment (SBA) and external examinations*) and V (*Students' preference between external exams and SBA*). Findings from section I is presented by means of independent-samples t-test to grasp the differences among the male and female students' self-reported responses. On the other hand, findings from section V is presented by means of measure of central tendency. The following tables statistically illustrate the students' responses:

I. Students' preference between external exams and SBA

An independent-samples t-test was run to compare the gender differences on the students' views of SBA and external examination. Next, effect size was calculated

to provide an indication of the magnitude of the differences between the groups (not just whether the difference could have occurred by chance).

Data revealed that some of the items listed in Table 1 and Table 2 had significant differences between the two genders in the sampled school. The items which had significant differences are first presented followed by the insignificant ones.

There is a significant difference in the students' views about *Taking external exam is a valuable experience* for male (Mean: 1.00) and female (M: 1.24), with a t-value of 2.219 and $p=0.041$. The result also indicates that taking external examinations is a more valuable experience for the sampled female students than to their male counterparts. The magnitude of the differences in the means

is considered large, with $\eta^2=0.133$. With regards to *External examinations force students to study harder*, there is a significant difference in the students' views, i.e., male (M: 1.29), female (M: 1.76) with a t-value of 3.024 and $p=0.005$. The female students in this school appear to have the urge to study harder for external examinations in comparison to the males. The magnitude of the differences in the means is large, too, with $\eta^2=0.22$. In relation to *SBA forces students to study harder*, there is a significant difference too, i.e., male (M: 1.24), female (M: 1.65) with a t-value of 2.578 and $p=0.015$. The findings indicate that the female students have the disposition to study harder due to the existence of SBA compared with their male counterparts. The magnitude of the differences in the means

Table 1
Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Taking external exam is a valuable experience	Male	17	1.00	.000	.000
	Female	17	1.24	.437	.106
External examinations force students to study harder	Male	17	1.29	.470	.114
	Female	17	1.76	.437	.106
SBA forces students to study harder	Male	17	1.24	.437	.106
	Female	17	1.65	.493	.119
A student's score on external examination is a good indication of how well he/she will be able to apply what has been learned	Male	17	1.24	.437	.106
	Female	17	1.65	.493	.119
A student's score on SBA is a good indication of how well he/she will be able to apply what has been learned	Male	17	1.00	.000	.000
	Female	17	1.59	.507	.123
All students work hard to achieve their best in external examinations	Male	17	1.47	.514	.125
	Female	17	1.76	.437	.106
All students work hard to achieve their best in SBA	Male	17	1.35	.493	.119
	Female	17	1.65	.493	.119

a. t cannot be computed because the standard deviations of both groups are 0.

Table 2
Levene's Test for Equality of Variances

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Taking external exam is a valuable experience	41.086	.000	-2.219	32	.034	-.235	.106
			-2.219	16.000	.041	-.235	.106
External examinations force students to study harder	.573	.455	-3.024	32	.005	-.471	.156
			-3.024	31.838	.005	-.471	.156
SBA forces students to study harder	2.140	.153	-2.578	32	.015	-.412	.160
			-2.578	31.556	.015	-.412	.160
A student's score on external examination is a good indication of how well he/she will be able to apply what has been learned	2.140	.153	-2.578	32	.015	-.412	.160
			-2.578	31.556	.015	-.412	.160
A student's score on SBA is a good indication of how well he/she will be able to apply what has been learned	497.778	.000	-4.781	32	.000	-.588	.123
			-4.781	16.000	.000	-.588	.123
All students work hard to achieve their best in external examinations	5.976	.020	-1.796	32	.082	-.294	.164
			-1.796	31.189	.082	-.294	.164
All students work hard to achieve their best in SBA	.000	1.000	-1.741	32	.091	-.294	.169
			-1.741	32.000	.091	-.294	.169

is large, with eta squared=0.172. As for *A student's score on external examination is a good indication of how well he/she will be able to apply what has been learned*, there is a significant difference, i.e., male (M:1.24), female (M:1.65) with a t-value of 2.578 and $p=0.015$. Female students here see their score in their external examination as a good indication of how well they will be able to apply what has been learned compared with the male students. The magnitude of the differences in the means is large, with eta squared=0.172. The last item which had a significant difference is *A student's score on SBA is a good indication of how well he/she will be able to apply what has been learned*, i.e., male (M:1.00), female (M:1.59) with a t-value of 4.781 and $p=0.000$. This indicates that the female students see their score in SBA as a good indication of how well they will be able to apply what has been learned compared with the male students. The magnitude of the differences in the means is large, with eta squared=0.417.

Two items from the tables above had insignificant differences. For the item, *All students work hard to achieve their best in external examinations*, no significant difference is found in the students' self-reported responses with a t-value of 1.796 and $p=0.082$. This may imply that both female and male students felt that they work equally hard to achieve their best in external examinations. Finally, the last item which had insignificant difference is *All students work hard to achieve their best in SBA*. There is no significant difference in the students' views with a t-value of 1.741 and

$p=0.091$. This may imply that both female and male students felt that they work equally hard to achieve their best in SBA

II. Challenges of Implementing School-based Assessment (SBA)

Measure of central tendency and standard deviation were computed to summarise data for students' perceived challenges of SBA.

As shown in Table 3, the mean value represents the average score of how the respondents perceived the challenges of implementing SBA in their school whereas the mode indicates the number of frequently chosen responses by the students. The five items (challenges) as shown in Table 3 are ranked from 1 (strongly disagree) to 4 (strongly agree).

The mean value for "*Our school does not have sufficient books on SBA*" is 2.94, which indicates that students on average agreed that the school does not have sufficient materials on SBA which seem to hinder the implementation of SBA in their school. On the other hand, the mean value for "*We find it difficult to understand the content of the SBA books*" is 3.24, which signifies that the students on average think that it is very difficult to understand the content of the SBA materials. The mean value for the item "*Teachers have the knowledge and skills to implement SBA*" is 1.56, which shows that the students on average slightly disagreed that their teachers have the knowledge and skills to carry out SBA tasks.

The mean value for the item, "*Students may not trust teachers' assessment in SBA*" is 3.06, which shows that on average,

respondents agreed that they may not trust the grades given by their own teachers in relation to SBA. The mode for this item is 4, which indicates that majority of the respondents strongly agreed with this statement.

The mean value for item “*We do not have adequate class time for carrying out SBA tasks*” is 3.26, which shows that respondents agreed and strongly agreed that they do not have adequate class time for carrying out SBA tasks. The mode value is 3, which indicates that majority of the respondents agreed that the class time allocated for SBA tasks is insufficient.

As shown in Table 3, the standard deviation of the item “Our school does not have sufficient books on SBA” is 1.071, which means data points are spread out over a wider range of values. Since the mean is 2.94 and the standard deviation is 1.071, it is estimated that approximately 95% of the scores will fall in the range of 2.94- (2*1.071) to 2.94+ (2*1.071), or between 0.798 and 5.082. The standard deviation of item “We find it difficult to understand the content of the SBA books” is 0.955, standard deviation for item “Teachers have

the knowledge and skills to implement SBA” is 0.894, standard deviation for item “Students and teachers may not trust teachers’ assessment in SBA” is 0.919 and standard deviation for item “We do not have adequate class time for carrying out SBA tasks” is 0.666.

DISCUSSION

The analysis of items for the first construct revealed that overall, ambivalent attitude was evident among male and female students in the sampled school. They were uncertain of which assessment method was more valuable for learning purposes. Moreover, some of them provided contradictory responses when asked about the two different assessment methods namely external examination and school-based assessment. Given such a reaction among this group of students and the MOE’s predilections for a ‘*synergistic assessment system*’ which combines both school assessment (*formative and summative*) and central assessment (*summative*), the question of which particular component has drawn serious attention among the student population in other schools and by extension,

Table 3
Challenges of Implementing School-based Assessment in our school

	Our school does not have sufficient books on SBA	We find it difficult to understand the content of the SBA books	Teachers have the knowledge and skills to implement SBA	Students may not trust teachers’ assessment in SBA	We do not have adequate class time for carrying out SBA tasks
Mean	2.94	3.24	1.56	3.06	3.26
Std. Deviation	1.071	.955	.894	.919	.666
Mode	4	4	1	4	3

states, then arises. Ideally, both components should be equally favoured by teachers and students alike as they complement each other (*intended washback*). Also, do the ministry officials (*macro-level stakeholders*) and the students (*micro-level stakeholders*) think along the same lines is one big question that has yet to be answered. By means of the quantitative data analysis of this study, it is safe to assume that the intended washback of the SBA on students was not fully achieved as there was a mismatch between what was required by the policy makers and the concept of SBA being operationalised by this group of students who participated in this study. In this circumstance, it is rather difficult to say that the SBA has negatively affected their knowledge and readiness as the researchers could not tell if the students were indifferent to the implementation of SBA or other factors contributed to their actions.

As can be seen in Table 3, students at the sampled school provided negative responses in relation to the items under the second construct. These students' account indicate that their schools were devoid of resources (materials) and those at their disposal were not helpful. One interesting dimension of the new assessment system is the teachers are now empowered to teach and assess their own students. However, judging the students' responses in this school, they appear to be quite pessimistic about this idea. Also, time constraint as a barrier to carry out SBA activities was perceived by these students. The students' responses for the second construct overall indicate that

there were barriers (materials, teachers' assessment literacy and time-constraint) in implementing SBA at the national schools around the country.

Overall, the students' ambivalence towards SBA has led the researchers to conclude that respondents of the study might not be very clear of the purposes, requirements and the potential benefits of the newly introduced school-based assessment. Moreover, the background of the respondents may have had an impact on their perceptions. Different language medium and academic achievement of students may also have impacted the way of learning and students' perceptions on assessment. Hence, the intended washback at this point in time was still at the surface level due to many uncertainties. The overarching examination-orientedness could inevitably be one of them.

CONCLUSION

For accurate interpretations of the results derived from this study, it is obligatory on the researchers' part to acknowledge its limitations. The first limitation of this study lies in the instrumentation employed to collect data from the participants. When this study was carried out, the teachers and other administrative officials in schools were not approachable due to various uncertainties with regards to the newly-introduced school-based assessment. Hence, students, who are one of the mostly affected stakeholders, were approached instead.

The researchers could only employ quantitative method i.e., the questionnaire

and therefore, no methodological triangulation (interviews and classroom observations) and data triangulation (teachers', policymakers' and parents' perspectives) could be done to triangulate the responses provided by the students in their self-reported questionnaires. Most importantly, classroom observation which is one of the most important elements which has been advocated by Alderson & Wall (1993) in washback studies to triangulate the claims made, unfortunately, could not be included in this study. Given the worldwide movement to combine *assessment of learning* with *assessment for learning*, a more comprehensive study which considers both teachers' and students' perspectives at possibly one of the regions or even nationwide in Malaysia may provide a holistic picture of the degree of impact observed. Second, the size of the sampled population is small in which the respondents were from one middle-banding school in one of the 14 states in Malaysia. Hence, generalising the findings to other contexts should be done with caution.

Notwithstanding the limitations, this study is possibly the first ever undertaken to gauge the students' perspectives of SBA in Malaysia. However, considering the fact that the sampled population of this study was the first batch of students to experience this new assessment and their perceptions and attitudes being shaped by other stakeholders at micro level (teachers and peers) and macro level (parents, policymakers and other stakeholders), the SBA has only managed

to produce a minimal washback effect (positive) among the sampled population. The researchers deemed that it is significant that the policymakers at national level should make every effort to ensure that all the necessary information in relation to the new assessment reaches the students without failure. Thus, this small scale study, if not clearly, has at least provided some insights on what kind of washback was observed in the selected school. Therefore, it can be used as a baseline study for further investigations of the impact of SBA on a bigger scale in Malaysia.

REFERENCES

- Alderson, J. C., & Wall, D., (1993). Does washback exist? *Applied linguistics*, 14(2), 115-129.
- Andrews, S. (2004). Washback and curriculum innovation. In Cheng, L., Watanabe, Y., & Curtis, A. (Eds.), *Washback in language testing: Research contexts and methods* (pp.37-50). New Jersey: Lawrence Erlbaum Associates.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3), 257-279.
- Bailey, K. M. (1999). *Washback in language testing*. TOEFL Monograph 15. Princeton, NJ Educational Testing Service.
- Baidzawi, A. M., & Abu. A. G. (2013). A comparative analysis of primary and secondary school teachers in the implementation of school-based assessment. *Malaysian Journal of Research*, 1(1), 28-36.
- Berry, R. (2011). Assessment Reforms Around the World. In Berry, R. & Adamson, B. (Eds.), *Assessment Reform in Education* (Vol. 14, pp. 89-102): Springer Netherlands.

- Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education: the United Kingdom.
- Blueprint, M. E. Blueprint 2013-2025. (2013). *Preliminary Report. Preschool to Post-Secondary Education*. Ministry of Education Malaysia.
- Chan, Y. F., Gurnam, K.S., & Rizal, Y., (2009). *School-Based Assessment Enhancing Knowledge and Best Practices*. Shah Alam: University Publication Centre (UPENA).
- Cheng, L., Watanabe, Y. & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. (1988). The t test for means. *Statistical power analysis for the behavioral sciences*, pp.19-74.
- Faizah, A. M. (2011). *School-Based Assessment in Malaysian Schools: The Concerns of the English Teachers*. Online Submission.
- Fredriksen, J., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Hamp-Lyons, L. (2000). Social, Professional and Individual Responsibility in Language Testing. *System*, 28(4), 579-591.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. (Unpublished manuscript). University of Reading.
- Khaniya, T. R. (1990). *Examinations as instruments for educational change: Investigating the washback effect of the Nepalese English exams*. (Unpublished doctoral dissertation). University of Edinburgh.
- McNamara, T. (2010). *Language Testing*, Oxford: Oxford University Press.
- Morrow, K. (1986). *The evaluation of tests on communicative performance*. Portal (Ed). Windsor, Berkshire: NFER/Nelson.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3) 241-256.
- MES (2014). *Information on School-based Assessment: Introduction to the school-based assessment*. Malaysia: Malaysian Examinations Syndicate, Ministry of Education.
- Nair, G. K. S., Setia, R., Samad, N. Z. A., Zahri, R. N. H. B. R., Luqman, A., Vadeveloo, T., & Ngah, H. C. (2014). Teachers' Knowledge and Issues in the Implementation of School-Based Assessment: A Case of Schools in Terengganu. *Asian Social Science*, 10(3), 186.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68(9), 679-682.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high stakes test. *Language Testing*, 22(2), 142-173.
- Saw, L. O. (2010). Assessment profile of Malaysia: high-stakes external examinations dominate. *Assessment in Education: Principles, Policy and Practice*, 17(1), 91-103.
- Shih, C-M. (2009). How Tests Change Teaching: A Model for Reference. *English Teaching: Practice and Critique*, 8(2), 188-206.
- Shohamy, E. (1991). International Perspectives of Foreign Language Testing Systems and Policy. In Ervin, G. (Ed.), *International perspectives on Foreign Language Teaching*, pp. 91-107. Lincolnwood, Ill.: National Textbook Company.
- Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: a study investigating the effects of test status on students' test preparation behaviours*. (Unpublished Doctoral dissertation). The Hong Kong Polytechnic University.

- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13(3), 334-354.
- Wall, D. (1997). Impact and washback in language testing. *Encyclopedia of language and education*, 7, 291-302.
- Yu, Ying. (2010). *The washback effects of school-based assessment on teaching and learning: A case study*. (Unpublished PhD dissertation). The University of Hong Kong.