

Examining the Factors Mediating the Intended Washback of the English Language school-based Assessment: Pre-service ESL Teachers' Accounts

Alla Baksh Mohd Ayub Khan^{1*}, Mohd Sallehudin Abd Aziz² and Siti Hamin Stapa²

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

²School of Languages and Linguistics, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ABSTRACT

This paper reports and discusses the quantitative findings of the perceived washback effects of the English language school-based assessment on teachers at the lower-secondary level in Malaysia. It aims to do so from the point of view of a group of pre-service English language teachers so as to provide both pre-service and in-service teachers with a better idea of the roles they need to play and the decisions they need to make about washback. Test washback studies reported in the literature, upon confirming the presence or absence of it, they have attempted to identify the factors within their respective contexts that appear to mediate the intended washback. Considering the recent implementation of the English language school-based assessment at the lower-secondary level intended to minimize the exam-orientedness within the Malaysian education system, the present study was therefore carried out to assess the mediation of factors at both macro and micro levels. The perspectives of a group of

pre-service English language teachers were looked into. Specifically, the study looked into the extent to which the perspectives of the teacher-trainees in a teacher training college in the country were in line/in conflict with that of the Malaysian Examinations Syndicate's (MES). Drawing mainly on the data from the survey administered, it was found that these teachers perceived there

ARTICLE INFO

Article history:

Received: 28 November 2017

Accepted: 11 July 2018

Published: 25 March 2019

E-mail addresses:

allabaksh@usm.my (Alla Baksh Mohd Ayub Khan)

salleh@ukm.edu.my (Mohd Sallehudin Abd Aziz)

sitihami@ukm.edu.my (Siti Hamin Stapa)

* Corresponding author

were some barriers at both the micro and macro levels, which inhibited the intended washback effect. Findings of the study may help policymakers in minimizing the barriers, which may inhibit the intended washback effect.

Keywords: Mediating factors, pre-service teachers, school-based assessment, test washback

INTRODUCTION

Globally, examinations have been widely used by educational authorities as tools/vehicles of policies to positively influence teaching and learning activities (Brindley, 2008; Shohamy, 2001). Upon closely scrutinising the tests/examinations that have been used for such purposes, it can be learnt that they are paper-and-pencil tests or other performance assessment, which are of high-stakes, mostly administered at the end of teaching and learning processes (*i.e.*, *summative-oriented*) and they serve various purposes such as proficiency, and achievement. (Barootchi & Keshvarz, 2002; Tsagari, 2004). A comprehensive review of literature in both general and language education has revealed that the educational authorities' disposition to use examinations for such purposes has provoked a succession of claims and counter-claims among researchers. For instance, Pearson (1988) had argued that tests/examinations could be used as *lever for change*. Popham (1987) had extended a similar notion with the term of *measurement-driven instruction* (MDI). Taken together, the above-mentioned terms

and others such as *test-curriculum alignment* (Shepard, 1990), *backwash* (Hughes, 1993) and *consequences* (Cizek, 2001) imply that what is tested is what gets taught and learnt in classrooms. In contrast to the above-mentioned terms widely used in the area of general education, washback or backwash in the area of language education is broadly defined as tests influencing teaching and learning (Alderson & Wall, 1993; Hughes, 1989). Upon reviewing its origin, it was learnt that discussions in relation to tests influencing teaching and learning were observed earlier in general education circles as it was only in the late 1980s, scholars in language education began looking into it (Cheng, 2008). In addition, a close scrutiny into its progression revealed that two schools of thought have been recorded, namely the psychometric and the social aspects of the washback phenomenon (Alderson & Wall, 1993; Bailey, 1996; Messick, 1996; Tsagari, 2007). The former propagates that good teaching and learning activities *i.e.*, positive washback can be created by means of the test design. On the contrary, the latter propagates that while validity is a property of a test, other factors/forces from both micro (*classrooms*) and macro (*education systems and societies*) levels may mediate the intended washback *i.e.*, what is intended by the test developers. The present study relied on the second school of thought *i.e.*, the social aspect of the washback effect.

It is worthy of note that critics from both general and language education circles have raised concerns about the usefulness of examinations as the primary measure

of learners' achievement. The reasons for which the critics disapproved of the conventional testing methods are, inter alia, the incapability of conventional testing methods (*summative-oriented*) to provide rich information to assess both processes (*growth*) and products (*achievement*) of learning, stakes which are attached to tests being relatively higher and thus adversely affecting both teachers and learners on curricular, educational and psychological levels, and the problems associated with teacher-made tests, which may potentially focus more on grading than on learning purposes. Alternatively, they call for the integration of 'alternative assessment' (*formative*) along with the conventional testing methods (*summative*) to ensure more reliable and valid measure of learning outcomes (Tsagari, 2007). In response to such concerns, education systems around the world have undertaken assessment reforms within their own contexts. Upon reviewing such reforms, Berry (2011) discovered that in both western and eastern parts of the world, a clarion call was made for education systems to implement both formative and summative assessment, to which education systems had responded in different ways. Specifically, the responses ranged from a total abolition of high-stakes testing in some settings to attempts to strike a balance between classroom and large-scale assessment in a synergistic system. The present context, in which the study was carried out i.e., Malaysia, there is a long history of summative examination-oriented system, in which four major public

examinations have to be taken by students at both primary and secondary levels (Fook et al. 2009; Khan et al., 2016; Lee, 2006; Ong, 2010; Pandian, 2002). 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., which the present study looked into, was initially known as *PMR (*Peperiksaan Menengah Rendah*) before it was renamed as Pentaksiran Tingkatan 3 (*PT3*) or Form 3 assessment in the year of 2014, is 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*) (Khan et al., 2016). With such a system, there was almost no room for the integration of formative assessment in classrooms. However, the government realized the deleterious effects of the high-stakes tests at the primary and lower-secondary levels, which made students to engage more in rote-learning and memorization at the expense of meaningful internalization (Khan & Aziz, 2015). 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. 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 (Malaysian

Examinations Syndicate, 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.

As it was highlighted earlier, theoretically, the present study relied on the social aspect of the washback phenomenon. Therefore, it has attempted to examine the factors from both macro and micro levels that appear to mediate the intended washback of the English language school-based assessment at the PT3 level. In this regard, the perceptions and attitudes of a group of pre-service English language teachers in the state of Penang, Malaysia were looked into. In the following sections, two bodies of literature i.e., school-based assessment and the washback effects, which concern the present study are reviewed.

School-Based Assessment (SBA)

The way scholars have defined SBA has not been consistent, as it was revealed by the literature. Related studies also indicated that the British writers refer to assessments, which are handled by teachers as formative assessment or teacher assessment whereas similar assessments are referred to as classroom assessment by the writers in the U.S. However, it has

to be noted that such assessment in the context of the U.S. involve a summative-orientation i.e., grading whereas it is not the case in the context of the UK (Brookhart, 2004). Owing to the long-standing history of summative-oriented examinations for accountability and selection purposes in the context of Asia, and the externally mandated examinations have been administered in many centralised education systems in the region, the relatively new paradigm of implementing assessment internally by means of empowering teachers in schools is referred to as SBA.

Upon looking into the definitions of SBA, it was found that Yussufu (1994) defined it as an assessment which involved the cumulative teacher judgment in relation to individual learner's work deriving from a systematic collection of grades or marks. He further highlighted that such an assessment was capable of serving as a monitoring instrument, which might provide diagnostic information to both teachers and learners to adjust their teaching and learning respectively. On the other hand, Ahmed and Williams (1994) defined it as an instrument, by means of which a wide range of assessment tasks and skills were assessed, had flexibility in the form of assessment i.e., written, and oral, and it employed open-ended questions. In addition, Izard (2001) as well as Raivoce and Pongi (2001) explained that SBA was often perceived as the process put in place to collect evidence of what students had achieved, especially in important learning outcomes that did not easily lend themselves to be assessed by the

paper-and-pencil tests. From the definitions provided in the foregoing, it can be seen that having teachers carrying out the internal (*schools*) assessment either on their own or with the assistance of the test developers, depending on the context in which it is implemented has been known under an assortment of names such as classroom assessment, teacher assessment and SBA. In the context of Malaysia particularly, SBA is a form of assessment conducted in schools, and is planned, administered, scored and reported in a mannered way based on the procedures and the guidelines provided by the Malaysian Examinations Syndicate (2011).

Figure 1 illustrates the English language school-based assessment implemented at the PT3 level in Malaysia.

According to Figure 1, there are two components under the PT3 English language

assessment. They are the central assessment and school assessment. The scope of these two assessment components is the national curriculum (i.e., KBSM) and they assess all four language skills. The combination of the two assessments is to complement each other in measuring the learning outcomes more reliably and validly.

The central assessment component is summative-oriented or assessment of learning at the end of lower-secondary education. The purpose is to assess learners' achievement at a particular time using instruments of comparable standards. MES prepares the instruments, scoring rubrics and guidelines while teachers administer, score and report the results. Also, listening and speaking (test) components included in this central assessment set it apart from the previous standardised exam (PMR) which was a traditional paper-and-pencil test.

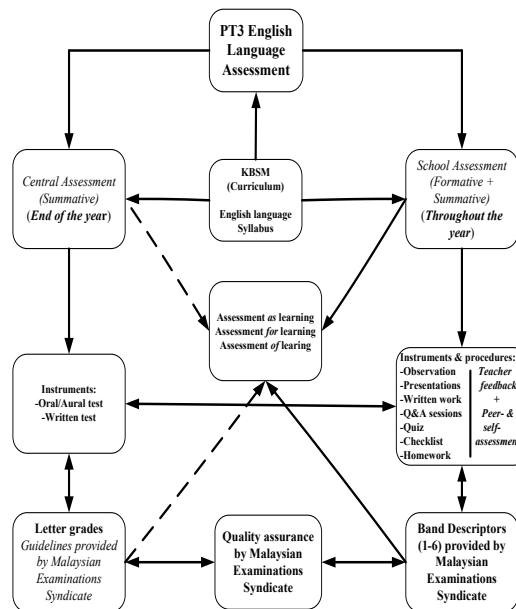


Figure 1. PT3 English language school-based assessment

The school assessment however looks at learners' learning progress and development. It is fully administered by teachers during teaching and learning, can be carried out formatively or summatively. The main focus of school assessment is to assess the learning that has taken place i.e., the acquisition of knowledge, skills and values in every subject learnt. Teachers are required to assess their students against a set of criteria (criterion-referenced assessment): a band scale provided by the MES. The bands range from 1 to 6 where 1 indicates the lowest level whereas 6 indicates the advanced level of learning the language.

Social Aspect of the Washback Effect

Some scholars (Morrow, 1986; Frederiksen & Collins, 1989; Khaniya, 1990) from the circles of both general and language education have widely asserted the existence of washback by heavily relying on the psychometric aspect of it without providing any empirical evidences. In language testing domain, this phenomenon came into prominence in early 1990s when Alderson & Wall (1993) disputed the assertions that a good test would produce beneficial teaching and learning (*positive washback*) and vice versa. In response, Alderson and Wall (1993) argued that a test by itself might not be the reason for the kind of teaching and learning observed in language classrooms as there might be other factors/forces within classrooms, schools, educational systems and society at work, which might hinder washback from

happening. They subsequently proposed 15 washback hypotheses in their seminal paper "Does washback exist?" which dealt with 'whats' and 'whos' were affected by tests. The 'whats' according to them were teaching - rate, sequence, degree and depth of teaching, and, learning - rate, sequence, degree and depth of learning and the 'whos' were 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), synthesized both Alderson and Wall's (1993) and Hughes' (1993) insights and presented the ideas with an addition of 'researchers' into the participants' category in the form of a diagram.

Alderson and Wall (1993) upon raising their arguments about the assertions made about washback effect, they themselves put their arguments to test in a washback study carried out in the context of Sri Lanka or better known as the Sri Lankan Impact study (Wall & Alderson, 1993) in which, they discovered that other than the test itself, there were other factors from both micro and macro levels, which mediated the intended washback effect. This was further confirmed by various other scholars from around the globe (Cheng, 1997; Cheng, 2005; Shohamy et al., 1996; Watanabe, 1996; Yu, 2010). Some of the salient findings of such studies are discussed in the following section.

Mediating Factors

The factors which have been identified by the empirical studies as influential in affecting washback to date are quite a number. They can be classified into four main categories: the teacher, resources, the school and the exam itself (Spratt, 2005). In relation to the teacher factor, four attributes have been cited in various studies i.e., their beliefs, their attitudes, their educational level and experience, and their personalities (Watanabe, 1996). In addition, resources are also said to be another factor which mediates the washback effect (Shohamy et al., 1996). Among the factors, which have frequently been mentioned are customized materials and exam support materials, such as exam specifications (Shohamy et al., 1996; Watanabe, 2000) and the types of textbooks available (Cheng, 1997; Hamp-Lyons, 1998). As for schools being a factor of washback, its atmosphere and cultural factors such as learning traditions (Watanabe, 2000); how much of pressure are put on teachers by the administrators to achieve results (Smith, 1991; Shohamy et al., 1996). Studies have also indicated that factors related to an exam/assessment itself may mediate the washback effect. Such factors may include its proximity, its stakes, the status of the language being tested, its purpose, the formats it employs (Shohamy et al., 1996), the weighting of individual papers (Lam, 1994), when the exam was introduced and how familiar it is to teachers (Andrews et al., 2002).

METHOD

Participants

The participants of this study were teacher-trainees at one of the teacher training colleges, in the state of Penang. Due to reasons of confidentiality, the sampled teacher training college preferred being anonymous. Therefore, the researchers agreed not to mention its name throughout the study. In addition, none of the respondents' personal details such as their names and ethnicity are revealed here. A total of 32 female and 6 male ($n = 38$) teacher trainees in the sampled teacher training college agreed to respond to the survey. When this study was carried out, it was already four years into the implementation of the English language school-based assessment at the lower-secondary level of education. Owing to the lack of empirical studies looking into the knowledge and skills of assessment among pre-service teachers in Malaysia, along with the inaccessibility for researchers to assess if assessment matters are equally stressed as part of the training, it was deemed necessary to investigate the mediating factors, which these prospective teachers perceive may appear to affect their teaching and assessing practices.

Instruments

It is of utmost importance to know if teaching and learning activities in classrooms are in line with what is expected of teachers and students especially in centralised education systems, in which decisions are made by policymakers at the

ministerial levels. In the case of the present study, as it was only focused on teachers' perspective, the researchers have referred to official documents which were issued by the Malaysian Examinations Syndicate (MES) to identify what is expected of teachers at the PT3 (*lower-secondary*) level. The researchers, upon identifying the intended washback, administered a survey which required the respondents ($n = 38$) to record their self-reported responses. They triangulated the responses provided by the pre-service teachers to see the extent to which these pre-service teachers' perspectives were in line or in conflict with that of MES'. Specifically, some mediating factors from both micro (*classroom*) and macro (*education system and society*) levels were included in the survey. These mediating factors in this study were operationalized as the challenges (*i.e., micro and macro levels*) from the perspective of pre-service teachers.

A validated questionnaire by Yu (2010), who conducted a mixed-methods case study on the washback effects of school-based performance assessment on

teaching practices among English language teachers in a Hong Kong secondary school, was adapted by this study. The original questionnaire was designed on a 6-point Likert scale, but the researchers had to transform it into a 5-point Likert scale. Therefore, the respondents were required to respond on a 5-point Likert scale ranging from 1 which indicates *Strongly disagree* to a score of 5 which indicates *Strongly agree*. Altogether, there are five sections in the questionnaire (Table 1).

The first test which is usually carried out on the data is the reliability test on the multi-item instruments used in research. The Cronbach's Alpha value is used to test the reliability of the items measuring each variable. While different levels of reliability are required, depending on the nature and purpose of the scale, Nunally (1978) recommended a minimum level of 0.7. It is a reliability measure coefficient that reflects how well items in a set are positively correlated to one another. The adapted questionnaire was revalidated by two local experts in the area of language testing and

Table 1
Sections of questionnaire and Cronbach's Alpha

Section	No. of items	Number of items Discarded	Cronbach's Alpha
I Perceptions on SBA	22	-	0.923
II Teaching Content	9	3	0.841
III Teaching Method	9	1	0.869
IV Assessment activities	9	2	0.846
V Challenges*	10	-	0.873

a master's degree holder from the area of Business management, and a reliability test was run for each section and for the entire instrument. An internal consistency test of the section V of the questionnaire revealed that its cronbach alpha value was at 0.873 (Table 1).

Considering the fact that the pre-service teachers sampled in the present study have yet to experience teaching in schools, the researchers felt that their responses for the first four sections (I, II III and IV) of the instrument may not be valid. Hence, only the results of section V of the survey are discussed in this paper. For the present study, after reviewing the related literature, the researchers were able to identify ten challenges which comprised both macro (*society and education system*) and micro levels (*classroom*). These challenges were mostly investigated in quite a number of washback studies (Wall & Alderson, 1993; Yu, 2010; Watanabe, 1996 & Cheng, 2005). Among the 10 challenges, four of them (i.e., *C1, C2, C4 and C10*) are challenges at the micro level whereas the rest (six) of them (*C3, C5, C6, C7, C8 and C9*) are challenges at the macro level. On reader-friendly grounds, the researchers have coded all the items as C1 till C10.

Following are the challenges at the micro level:

- C1 – students' current English level*
- C2 – Class size*
- C4 – Classrooms with students of mixed-abilities*

C10 – Students do not prefer being assessed by their own teachers; and

Following are the challenges at the macro level:

C3 – inadequate textbooks and other available teaching resources

C5 – the lack of teaching and learning aids and facilities

C6 – too heavy workload

C7 – inadequate time for carrying out the school assessment

C8 – lack of information from the ministry

C9 – parents do not trust the teachers' grades

Data Collection and Analysis

Upon obtaining the approval from the state education department of Penang to carry out this study, an acquaintance of the researchers who had been a senior staff member at the teacher training college was approached to assist the researchers with the data collection among the teacher-trainees. The survey forms were completed by the teacher-trainees under the senior staff's supervision with a return rate of 95%. The Software Package for Social Sciences (SPSS, V21) was used to analyse the data. The measure of central tendency was carried out to see the relationship between the relevant demographic data and the challenges (mediating factors) at both macro and micro levels from the perspective of the sampled pre-service teachers.

RESULTS

Demographic Profile

The relevant demographic data of the sampled pre-service teachers in this study are gender, age, language course qualifications, educational qualifications, experience in teaching the English language subject and optionist/non-optionist teachers. The data are presented descriptively i.e., in the form of frequency and percentage (Table 2).

Referring to Table 2, it can be noted that there is a stark contrast between the number of male and female teachers within the sampled teacher training college. This imbalanced gender gap is a reflection of the overall statistics of teacher-trainees enrolled at teacher training colleges and

even the statistics of in-service teachers at both primary and secondary schools nationwide. The tables below (Table 3, 4 and 5) indicate such imbalanced gender gaps observed at the enrolment rate in teacher training colleges and the teaching force of the country. These tables were cited from the official documents from the Ministry of Education (*quick facts* for the year of 2015).

In relation to age, 37 of the respondents were of 20-29 years old whereas only one respondent was of 30-39 years old. As for the language courses qualification, 36 of them reportedly had SPM level English language qualifications. As for the educational qualifications, majority of the respondents reported that they were at Bachelor's degree level. The researchers

Table 2
Relevant demographic data of pre-service teachers

Demographic Profile		Frequency	Percentage (%)
Gender	Male	6	15.8
	Female	32	84.2
Age	20-29	37	97.4
	30-39	1	2.6
Language Courses Qualification	SPM	36	94.7
	MUET	1	2.6
	IELTS	1	2.6
Educational Qualifications	Bachelor's Degree	35	92.1
	Doctorate (PhD/ EdD)	1	2.6
	Other	2	5.3
Experience in teaching English	less than 10 years	38	100.0
Did you opt to teach English as preferred subject?	Yes	38	100.0

Table 3
Enrolment of teacher trainees in institutes of teacher education

Institute of Teacher Education		Enrolment		
		Male	Female	Total
1.	PGM Kampus Perlis	368	624	992
2.	IPGM Kampus Darul Aman	437	938	1375
3.	IPGM Kampus Sultan Abdul Halim	409	1223	1632
4.	IPGM Kampus Pulau Pinang	282	827	1109
5.	IPGM Kampus Tuanku Bainun	391	1050	1441
6.	IPGM Kampus Ipoh	498	1341	1839
7.	IPGM Kampus Bahasa Melayu	282	875	1157
8.	IPGM Kampus Bahasa Antarabangsa	209	489	698
9.	IPGM Kampus Ilmu Khas	235	691	926
10.	IPGM Kampus Pendidikan Teknik	259	454	713
11.	IPGM Kampus Pendidikan Islam	419	947	1366
12.	IPGM Kampus Raja Melewar	457	1406	1863
13.	IPGM Kampus Perempuan Melayu	147	870	1017
14.	IPGM Kampus Tun Hussein Onn	450	1263	1713
15.	IPGM Kampus Temenggong Ibrahim	477	1370	1847
16.	IPGM Kampus Tengku Ampuan Afzan	370	961	1331
17.	IPGM Kampus Dato' Razali Ismail	375	945	1320
18.	IPGM Kampus Sultan Mizan	344	571	915
19.	IPGM Kampus Kota Baharu	514	969	1483
20.	IPGM Kampus Gaya	233	522	755
21.	IPGM Kampus Kent	195	485	680
22.	IPGM Kampus Tawau	182	271	453
23.	IPGM Kampus Keningau	198	344	542
24.	IPGM Kampus Batu Lintang	288	755	1043
25.	IPGM Kampus Sarawak	258	532	790
26.	IPGM Kampus Rajang	243	489	732
27.	IPGM Kampus Tun Abdul Razak	293	625	918
TOTAL		8813	21837	30650

Source: Institute of Teacher Education, MOE

Table 4

Number of in-service teachers at the primary level by gender

Age Group	<25	25-29	30-34	35-39	40-44	45-49	50-54	55-58	>58	Total
Male	1839	7748	13806	13604	11103	10922	7943	4402	831	72198
Female	3559	18117	36862	34546	28767	23701	17448	5642	697	241537
Total	5398	25865	25865	50668	48150	39860	34623	25391	10444	241537

Source: School Management Division, MOE (Data as of 31 May 2015)

Table 5

Number of in-service teachers at the secondary level by gender

Age Group	<25	25-29	30-34	35-39	40-44	45-49	50-54	55-58	>58	Total
Male	106	5375	8081	7856	8724	10747	9176	4200	593	54858
Female	440	16896	27961	21730	21730	20740	13438	4141	413	127616
Total	546	22271	36042	29713	30454	31487	22614	8341	1006	182474

Source: School Management Division, MOE (Data as of 31 May 2015)

would like to highlight that these teachers were undergoing their teacher training at their first degree level. Therefore, they have yet to earn their Bachelor's degree. However, it is considered their highest qualification because they were at their final stages of their training program when this study was carried out. It can be argued that these pre-service teachers may not have taught the English language in schools at all. However, it is noteworthy that these teachers have the component of practicum embedded within their teacher training programs, in which they are supposed to teach the English language in schools. To the knowledge of the researchers, teachers from teacher training programs undergo their practicum for an estimated 6 months. Thus, they fall under the category of teaching experience with less than 10 years as stated in the

instrument. As for opting to teach English, all of them opted to teach English as their preferred subject.

Means Comparison Test

The Means procedure is useful for both description and analysis of scale variables. The Means procedure calculates subgroup means and related univariate statistics for dependent variables within categories of one or more independent variables. The following tables present the means comparison in which the participants' (*teachers'*) mean score of selected items of their demographic profile (*i.e., gender and experience of teaching English*) are compared against the mean score of perceived challenges.

From Table 6, we can see that overall, fairly high mean scores (> 3.5) are recorded for almost all the challenges. However,

it is quite interesting to note that of the total ten challenges (C1 – C10), the highest mean scores were observed at C4. Notwithstanding, the mean score of C4 for male participants is not the highest, unlike the female participants. The highest mean score for male participants was observed at C1 (4.67). The second highest mean score was recorded at C7 (3.97). However, it is noteworthy that it is also the highest mean score for the female participants but not for the male participants. On the contrary, the second highest mean score for male participants was recorded at C3 and C4 (3.83). Interestingly, the responses provided by the male respondents considerably contradicted with their female counterparts. Specifically, the male participants did not

perceive C5, C6, C9 and C10 as significant challenges whereas their female counterparts perceived almost every challenge as a significant one.

Referring to Table 7, it can be noted that all the participants sampled in this study had less than 10 years of teaching experience. It was earlier discussed in this paper as to why these teachers had less than 10 years of teaching experience. It is noteworthy that almost every challenge was rated as significantly challenging by the participants of the study. The highest mean score was recorded at C4 (4.05). The second highest mean score was recorded at C7 (3.97) and the lowest mean score was recorded at C2 (3.50).

Table 6
Means comparison (gender & perceived challenges)

Gender		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Male	Mean	4.67	3.50	3.83	3.83	3.00	3.33	3.50	3.50	2.83	3.00
	N	6	6	6	6	6	6	6	6	6	6
	Std. Deviation	0.516	0.837	1.169	0.753	1.414	1.033	1.643	1.761	1.472	1.549
Female	Mean	3.63	3.50	3.78	4.09	3.97	3.97	4.06	3.81	3.81	3.62
	N	32	32	32	32	32	32	32	32	32	32
	Std. Deviation	0.907	0.842	0.706	0.689	0.822	0.861	0.840	0.859	0.780	0.907
Total	Mean	3.79	3.50	3.79	4.05	3.82	3.87	3.97	3.76	3.66	3.53
	N	38	38	38	38	38	38	38	38	38	38
	Std. Deviation	0.935	0.830	0.777	0.695	0.982	0.906	1.000	1.025	0.966	1.033

Table 7

Means comparison (English teaching experience & perceived challenges)

English Teaching Experience		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Less than 10 years	Mean	3.79	3.50	3.79	4.05	3.82	3.87	3.97	3.76	3.66	3.53
	N	38	38	38	38	38	38	38	38	38	38
	Std. Deviation	0.935	0.830	0.777	0.695	0.982	0.906	1.000	1.025	0.966	1.033
Total	Mean	3.79	3.50	3.79	4.05	3.82	3.87	3.97	3.76	3.66	3.53
	N	38	38	38	38	38	38	38	38	38	38
	Std. Deviation	0.935	0.830	0.777	0.695	0.982	0.906	1.000	1.025	0.966	1.033

DISCUSSION

From the foregoing results, we can deduce that the sampled pre-service teachers have unanimously raised the issue of one particular micro level challenge i.e., classrooms with students of mixed-abilities (C4) as the most significant one for them. Interestingly, for both categories of gender and teaching experience, the pre-service teachers responded that this micro level challenge is of the most significant. The researchers feel that the sampled pre-service teachers may not have been equipped with sufficient knowledge base in relation to formative assessment which is about diagnosing students' weaknesses and using the test-/assessment-derived information as feedback and feedforward to help improve their own teaching and helping their students at the same time by means of peer- and self-assessment. In addition, it is also felt that these teachers may lack innovativeness in relation to mixing low-performing, mediocre and high-performing students

within classrooms to make teaching and learning activities more meaningful and fun. The researchers are also of the view that these teachers may have the disposition to focus more on individual activities rather than doing group activities, discussions, etc., which may help deal with the micro level issue they have raised. Therefore, the pre-service teachers' perspectives in relation to classrooms with students of mixed-abilities are in conflict with that of MES'.

Moving on, inadequate time for carrying out the school assessment (C7) i.e., one of the challenges at the macro level, recorded the second highest mean score for both gender and teaching experience categories respectively. It is again the researchers' view that the sampled pre-service teachers may not realize that the school-based assessment which is presently implemented at the lower-secondary level is a synergistic one in which, the central assessment which comes at the end of the year and the ongoing school assessment complement each other. More importantly, the school assessment is

considered low-stakes as it involves various classroom activities (Figure 1) throughout the years of form 1, 2 and 3. It does not directly involve the MES in assessing the students' language learning outcomes due to the recent decentralization of decisions in relation to assessment activities. The main focus for the school assessment component should be on the growth or the process of language learning over a period of time rather than the products (*i.e.*, *grades*). The researchers also wonder if these teachers are aware that the PT3 test battery itself is considered low-stakes as it does not involve any certification. In other words, it does not involve any serious implications on the students' lives except for moving on to the next form *i.e.*, form 4. These teachers' response in relation to this macro level challenge indicates that the overarching exam-orientedness which the MoE wants to minimize, after four years into the implementation of English language school-based assessment, have yet to taper off.

The male pre-service teachers' responses were to some extent, contradictory to those of their female counterparts'. Specifically, they felt that the lack of teaching, learning aids and facilities (C5) and parents did not trust teachers' grades (C9) *i.e.*, challenges at the macro level were not real challenges for them. It may indicate that these particular male teachers may have been creative and innovative, and acted on their own initiative to supplement the limited resources available and their experiences indicated that parents whom they encountered do indeed trust the grades

assigned by teachers. Notwithstanding, their female counterparts thought otherwise. Lastly, these respondents who have less than 10 years of English teaching experience felt that the micro level challenges of class size (C2) and students did not prefer being assessed by their own teachers (C10) were not significant. It implies that these teachers may have positive attitudes about teaching classes of considerably bigger sizes and their own experiences may have indicated that students do indeed prefer being assessed by their own teachers.

As the teachers sampled in the present study had yet to experience real teaching, their responses, which were found contradicting (between males and females for instance), may reflect the level of exposure to assessment matters. The researchers feel that pre-service teachers are required to be given more exposure to the synergistic assessment system that has been implemented in recent years as part of their training activities. Such training activities may serve as a foundation for them to pursue assessment matters better when they begin teaching in schools. More importantly, such an exposure may help create positive washback.

CONCLUSION

It is of utmost importance to highlight the limitations of the present study to help researchers who intend to pursue the issues raised in the present study accurately. The present study employed only a survey, by means of which, it investigated the contextual factors at both

micro and macro levels, which mediated the intended washback effects of the English language school-based assessment from the perspective of the sampled teacher-trainees. The researchers were not able to carry out methodological (interviews) and data triangulation (policymakers' and parents' perspectives) to triangulate the responses provided by the teacher-trainees in their self-reported questionnaires. However, it is noteworthy that classroom observations were not carried out in the present study as it only involved teacher-trainees who were not involved in regular teaching practices. Considering the respondents of the present study were sampled from one of the teacher training colleges in the northern region, generalizing the findings to other states and contexts should therefore be done with caution. Despite some of its limitations, the present study has shed some light on the challenges that these prospective teachers perceive they may encounter upon being posted to schools in the country.

To the knowledge of the researchers, not many studies to date have focused on studying the attitudes and behaviours of teacher trainees in relation to the English language school-based assessment. Therefore, more studies are needed to better understand their views, knowledge and understanding on this new assessment system which may culminate in helping policymakers making necessary adjustments in their policies and more importantly, teachers can be better trained while undergoing their training before being posted to schools. The success of training the teachers well by means of

strengthening their knowledge base and understanding may help the MoE spare financial incentives which may have to be spent for in-service training for teachers.

REFERENCES

- Ahmed, A., & Williams, H. (1994). *School-based assessment: Improving students in the process*. Paper presented at the 1993 IAEA Conference, Reduit, Mauritius.
- Alderson, C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14(2), 115-129.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting Washback—A Case-Study. *System*, 30(2), 207-223.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44(3), 279-288.
- Berry, R. (2011). Assessment reforms around the world. In R. Berry & B. Adamson (Eds.), *Assessment Reform in Education* (pp. 89-102). Dordrecht: Springer Netherlands.
- Brindley, G. (2008). Educational reform and language testing. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (pp. 2495-2508). New York, NY: Springer Science+Business Media.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.

- Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A Washback Study* (Vol. 21). Cambridge, the United Kingdom: Cambridge University Press.
- Cheng, L. (2008). Washback, Impact and Consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (pp. 2479-2494). New York, NY: Springer Science+Business Media.
- Cizek, G. J. (2001). More Unintended Consequences of High-stakes Testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Fook, C. Y., Sidhu, G. K., & Yunus, M. R. M. (2009). *School-based Assessment: Enhancing Knowledge and Best Practices*. Shah Alam, Malaysia: Pusat Penerbitan Universiti, Universiti Teknologi MARA.
- Frederiksen, J. R., & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher*, 18(9), 27-32.
- Hamp-Lyons, L. (1998). Ethical Test Preparation Practice: The Case of the TOEFL. *TESOL Quarterly*, 32(2), 329-337.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge, the United Kingdom: Cambridge University Press.
- Hughes, A. (1993) *Backwash and TOEFL 2000* (Unpublished manuscript). University of Reading, England.
- Izard, J. (2001). Implementing school-based assessment: Some successful recent approaches used in Australia and the Philippines. In *Proceedings of the First International Conference of the Association of Commonwealth Examinations and Accreditation Bodies*. Reudit: Mauritius Examinations Syndicate.
- Khan, A. B. M. A., & Aziz, M. S. A. (2015, December 30). Tail wagging the dog? *New Straits Times*. Retrieved May 30, 2017, from <https://www.nst.com.my/news/2015/12/119658/tail-wagging-dog>
- Khan, A. B. M. A., & Aziz, M. S. A., Tayeb, Y. A., & Norhaslinda, H. (2016). Washback Effect of School-based English Language Assessment: A Case-Study on Students' Perceptions. *Pertanika Journal of Social Sciences & Humanities*, 24(3), 1069 - 1086.
- Khaniya, T. R. (1990). *Examinations as instruments for educational change: Investigating the washback effect of the Nepalese English exams*. (Unpublished PhD thesis). University of Edinburgh, England.
- Lam, H. P. (1994). Methodology washback-an insider's view. In *Bringing About Change in Language Education: Proceedings of the International Language in Education Conference* (pp. 83-102). Hong Kong.
- Malaysian Examinations Syndicate. (2011). *Panduan dan Peraturan PBS* [School-based assessment guidelines and rules]. Putrajaya, Malaysia: Lembaga Peperiksaan.
- Lee, M. N. (2006). Centralized decentralization in Malaysian education. In C. Bjork (Ed.), *Educational Decentralization* (pp. 149-158). Netherlands: Springer.
- Malaysian Examinations Syndicate. (2014). *Panduan Pengurusan Pentaksiran Berasaskan Sekolah* [School Based Assessment Management Guide]. Putrajaya, Malaysia: Lembaga Peperiksaan.
- Messick, S. (1996). Validity and Washback in Language Testing. *ETS Research Report Series*, 1996(1), 1-18.
- Ministry of Education (MoE). (2015). *Quick Facts: Malaysian Educational Statistics 2015*. Kuala Lumpur: Educational Planning and Research Division, Ministry of Education, Malaysia.
- Morrow, K. (1986). The Evaluation of Tests of Communicative Performance. *Innovations in Language Testing*, 3, 1-13.
- Nunally, J. C., (1978). *Psychometric Theory* (2nd Ed). New York, USA: Mc Graw-Hill.

- Ong, L. S., (2010). Assessment profile of Malaysia: High-stakes External Examinations Dominate. *Assessment in Education: Principles, Policy & Practice*, 17(1), 91-103.
- Pandian, A. (2002). English Language Teaching in Malaysia Today. *Asia Pacific Journal of Education*, 22(2), 35-52.
- Pearson, I. (1988). Tests as Levers for Change. *ESP in the Classroom: Practice and Evaluation*, 128, 98-107.
- Popham, W. J. (1987). Preparing Policymakers for Standard Setting on High-stakes Tests. *Educational Evaluation and Policy Analysis*, 9(1), 77-82.
- Raiwoce, A., & Pongi, V. (2001). School Based Assessment: A First Hand Experience in the Small Island States of the South Pacific. In *Proceedings of the First international Conference of the Association of Commonwealth Examinations and Accreditation Bodies*. Reduit: Mauritius Examinations Syndicate.
- Shepard, L. A. (1990). Inflated Test Score Gains: Is the Problem Old Norms or Teaching the Test? *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Shohamy, E. (2001). Democratic Assessment as an Alternative. *Language Testing*, 18(4), 373-391.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test Impact Revisited: Washback Effect Over Time. *Language Testing*, 13(3), 298-317.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Tsagari, D. (2004). *Is there life beyond language testing? An introduction to alternative language assessment*. CRILE Working Papers No. 58 (2004). University of Lancaster: the United Kingdom.
- Tsagari, D. (2007). *Review of Washback in Language Testing: What Has Been Done? What More Needs Doing?* Retrieved May 30, 2017, from <https://files.eric.ed.gov/fulltext/ED497709.pdf>
- Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms. *Australian Review of Applied Linguistics. Supplement Series*, 13(1), 208-239.
- Watanabe, Y. (2000). Washback effects of the English section of the Japanese university entrance examinations on instruction in pre-college level EFL. *Language Testing Update*, 27, 42-47.
- Yu, Y. (2010). *The washback effects of school-based assessment on teaching and learning: A case study*. (HKU Theses Online) The University of Hong Kong, Hong Kong.
- Yussufu, A. (1994). *A model for using school-based assessment*. Paper presented in IAEA Conference, Reduit, Mauritius.