Language Testing in Focus: An International Journal

www.EUROKD.com

ISBN: 2717-9087

2023 (7)

Implementing formative assessment in Malaysia: Teachers' viewpoints

Alla Khan[1], Norhaslinda Hassan[2*], Azril Ali[3]
[1]*School of Languages, Literacies & Translation, Universiti Sains Malaysia, Malaysia*
[2]*Academy of Language Studies, Universiti Teknologi MARA, Malaysia*
[3]*School of Languages, Literacies & Translation, Universiti Sains Malaysia, Malaysia*

Keywords
Assessment as Learning, Washback, Peer- and Self-Assessment, Formative Assessment, Malaysia

Received
09 May 2022
Received in revised form
10 March 2023
Accepted
25 March 2023

Correspondence concerning this article should be addressed to: haslinda.hassan@uitm.edu.my

ABSTRACT

The paper aims to study the implementation of one of the dimensions of formative assessment i.e., assessment as learning. It has specifically attempted to investigate the practices of peer- and self-assessment among teachers in schools. Considering the recent growth in implementing formative assessment around the world; both in schools and higher learning institutions, this study aims to shed some light on the practices of it in schools, which can help policymakers adjust their policies in higher learning institutions. This study employed a quantitative method, in which a questionnaire is distributed among teachers in four states in the northern region of Malaysia. This paper offers insights into how teachers attempted to practise peer- and self-assessment in schools in the northern region of Malaysia. As this is a washback study, it not only reports the factors that were discovered, but it also attempts to report the factors mediating the washback. This study can be considered a contribution in terms of shaping the research into assessment as learning at its early stage.

Introduction

It is very common for educational policymakers to include new constructs in examinations. This is because examinations are believed to have the power to draw the attention of almost every stakeholder in education systems. In other words, examinations are the foci of

https://doi.org/10.32038/ltf.2023.07.03

policymakers when it comes to ensuring students (the ultimate stakeholders) pay attention to what the policymakers feel are important about the contents of a subject, course, etc. (Al Hinai & Al Jardani, 2021; Alqahtani, 2022; Brindley, 2008; Shohamy, 2001). Such a practice goes way back to the 1980s when some have argued that tests/examinations can be used as 'lever for change' (Pearson, 1988). Mostly, centralised education systems implement such policies possibly owing to the cascade method being used mostly (Shohamy, 1991). Upon a close scrutiny, the concerned examinations have serious consequences for teachers and students (high-stakes), summative in nature and serve multiple purposes such as proficiency, achievement, placement, etc. and they still reign supreme in many education systems (Butterfield et al., 1999; Tsagari, 2004). However, it is noteworthy that education systems are presently focusing more on striking a balance between classroom and large-scale assessment in a synergistic system (Berry, 2011).

With regard to the context of the present study, Malaysia has a long tradition of summative tests, in which students are required to take four main public exams both at primary and secondary levels (Alla Baksh et al., 2016). With such a framework, formative assessment in the classroom was essentially impossible to incorporate. Due to this, students may end up focusing more on memorization and rote learning at the price of meaningful internalisation. Hence, the government has embarked on assessment reforms at both the primary and secondary levels, in which the formative assessment is made a priority. It is also worth highlighting that in the year of 2002, a school-based oral assessment (SBOA) component was introduced for both Bahasa Malaysia and the English language subjects at the upper-secondary level (*year 10 and 11*) for which students are assessed both formatively and summatively by their own teachers by adhering to the directions offered by the Malaysian Examinations Syndicate (MES). However, the result of the English language SBOA (year 11) component does not have any bearing on the decisions made about students' future undertakings (i.e. low-stakes) as it is the summative component which carries much weighting that is taken into account for decisions such as the awarding of Public Service Department (PSD) or other scholarships and employment opportunities. In addition, some related studies (Majid et al., 2011) have also highlighted that the English language SBOA was not carried out as intended by the MES owing to teachers' lack of understanding. In other words, formative assessment has yet to be incorporated in the classroom of Malaysian schools. Although the recent assessment reform involved both the primary and lower-secondary levels, the present study focuses on the lower-secondary level only. In 2012, the Malaysian government introduced school-based assessment for all subjects offered at the lower-secondary level. Specifically, for the English language subject, compared to the previous system, there are presently two components, namely the central assessment (Form 3 assessment) and the school assessment, which the present study has looked into.

Literature Review

In relation to formative assessment, scholarly discussions in the literature reveal that teachers and learners are actively involved in the assessment process, in which they assess themselves and each other (Black et al., 2003; Dilova, 2021; Yan et al., 2021). As for teachers, it is about diagnosing or identifying their learners' strengths and weaknesses to offer constructive

feedback to close the gaps identified. Such practices may also guide the teachers to make necessary adjustments to their teaching methods, etc. In addition, it is also their duty to guide their learners to carry out both peer- and self-assessment activities, which are capable of promoting ownership and autonomy of learning among learners (Topping, 2003; Yu, 2010). In summary, teacher feedback, peer- and self-assessment are seen as the significant features of the formative assessment. Also, they are interrelated as the teachers' formative feedback may encourage learners to probe deeper into their strengths and weaknesses themselves. On the other hand, the meta-cognition of learners especially through self-assessment may encourage them to get in touch with their teachers (teacher feedback) and their peers (peer-assessment). The definitions, theoretical underpinnings and some rules of thumb for these constructs as advocated by scholars are briefly discussed in the following sections.

Feedback

Feedback in relation to assessment activities within education systems, which may affect students' learning can be broadly defined as "information provided by an agent (e.g. teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81). In line with the recent developments in many education systems, which focus more on formative assessment, feedback to students to improve their learning is more important. However, a comprehensive review of literature indicates that although feedback is considered as a significant factor for improving student learning, little is known about how students understand and experience feedback within the classroom settings (Harris et al., 2014). One of the pioneer studies focusing on AfL (Black & William, 1998) concluded that formative assessment is capable of establishing clear objectives for students, appropriate learning tasks, and giving them constructive feedback may enhance learning, especially for underachievers. In addition, in delving deeper into what can be considered as formative feedback and its mechanism, it has been argued that feedback from assessment can be considered formative when it highlights the quality of the task or learning process, reveals misconceptions, and helps learners create more productive learning strategies (Black & Wiliam 1998; Hattie & Timperley, 2007; Lee, 2006; Mok, 2012).

Peer-and-self-assessment

Peer-assessment can be defined as 'an arrangement in which individuals consider the amount, level, value, worth, quality of success of the products or outcomes of learning of peers of similar status' (Topping & Ehly, 1998, p. 250). Specifically, peer assessment is believed to aid students in developing abilities and skills that are unavailable to them in a learning environment where the teacher is the sole assessor. In other words, it gives students the opportunity to assume responsibility for analysing, monitoring, and assessing aspects of both the learning process and product of their peers. Self-assessment on the other hand can be defined as 'procedures by which the learners themselves evaluate their language skills and knowledge' (Bailey, 1999, p. 227).

As it was discussed earlier, formative assessment allows learners to be actively involved in the assessment process. This include helping learners plan their language learning, identify their strengths and weaknesses, identify areas for corrective actions/feedback and develop meta-

cognitive and other personal and professional transferable skills (Boud, 2000; Brown & Knight 1994; Davison, 2007). More interestingly, triangulating the formative feedback offered by teachers with both the self- and peer-assessment may minimize the potential threats to learning quality. In addition, valuable triangulation in the assessment process can be achieved by both peer- and self-assessment and if properly done, both can have measurable formative effects on learning (Topping, 2003). When we consider in what circumstances these assessment tools may serve well, some scholars have argued that in the event of students having to sit for a test which is made up of items that have only one correct answer, the traditional approach can be considered suitable. However, having a single assessor may result in potentially biased evaluations in performance tests, such as oral presentations, written compositions, and role-plays. Considering the formative assessment, in which various methods of assessment are involved, the limitations of teacher assessments can be dealt with (Matsuno, 2009). Therefore, in the field of language education, alternative assessments, i.e., self-assessments and peer-assessments have been the increasingly focused (Hargreaves et al., 2002).

A Panoramic Assessment of Washback

In the area of language assessment, scholars acknowledge the relationship between tests, teaching and learning although some differences were observed in relation to the terms used by them in describing such a relationship. Washback (Alderson & Wall, 1993) or backwash (Hughes, 1989) generally refers to the influence of tests on teaching and learning. This construct is rooted in the notion that tests or examinations can and should drive both teaching and learning and therefore, it is referred to as measurement-driven instruction (Popham, 1987). Such a goal can be achieved depending on the extent to which the content and format of both the examination and the curriculum match or overlap. In this regard, scholars widely believe that the closer the fit or match, the greater the potential improvement on the test (Cheng 2000). Such a scenario is described as curriculum alignment by Shepard (1993). It has to be noted that some scholars have considered such an alignment or matching of the examination and the curriculum as unethical and may threaten the validity of tests (Haladyna et al., 1991, p. 4; Widen et.al, 1997). Notwithstanding, such alignments have over the years been widely implemented in various education systems. Reviewing the literature may clearly indicate that either a new or revised examination is introduced into education systems with the aim of improving both teaching and learning. For instance, the assessment reforms in the context of Hong Kong (Cheng, 1998; Yu, 2010) and Malaysia (Tan, 2009).

Notably, the discussion in the area of language testing revolving around such a phenomenon has used three terminologies. They are ‘backwash’ (Hughes, 1989), ‘washback’ (Alderson & Wall, 1993) and ‘impact’ (Wall, 1997). ‘Impact’ is considered as a broader concept in the area of language testing, which Wall (1997, p. 291) defined it as “…any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole.” Saville (2009, p.3) defined impact as ‘..the superordinate concept covering the effects and consequences of tests and examinations throughout society, whereas washback is more limited and refers to the influence of tests and examinations in teaching and learning contexts (e.g. classrooms and schools)’. In addition, Bachman and Palmer (1996) in addressing the complex nature of this phenomenon suggested that impact may have to be

considered within social contexts, considering the variables such as social goals and values, the educational systems in which tests are used, and the potential outcomes of their uses. It is also worth highlighting that washback is a form of impact and the term is commonly used to refer specifically to the effects of tests on teaching and learning. Hamp-Lyons (1997) suggested a view of test influence, which falls between the narrow one of washback and the all-encompassing one of impact. It has to be noted that the terms 'washback' and 'backwash' have over the years been used interchangeably among language testing writers. In his attempt to explain the differences between these two terms, Alderson (2004, p. xi) said 'there is none'. Nevertheless, 'washback' has been the preferred term among the writers of British applied linguistics.

Considering the nomenclature of describing tests influencing teaching and learning along with the assessment reform in the present context, in which a synergistic English language school-based assessment was introduced, it is clear that the assessment reform is rooted in the notion that tests or in this context, the English language school-based assessment in a centralised and top-down education system can and should influence teaching and learning. In other words, the policymakers (e.g. the Malaysian Examinations Syndicate) expect both structural and attitudinal changes among both teachers and learners. In this regard, following the convention practised among language testing scholars, the present study uses the terms 'washback' and 'backwash' interchangeably in addressing the micro context i.e. classroom and its immediate surroundings (schools) and impact to address the influences on the macro levels (i.e. policymaking, publishers, private tuition centres and parents).

Watanabe (2004) outlined five dimensions of the washback phenomenon. They are washback specificity, washback intensity, washback length, washback intentionality and washback value. Among the five, the washback value i.e., promoting positive washback and ameliorating negative washback is the most relevant for the present study. The findings of the washback studies to date have indicated that the washback from assessment can be positive or negative (Alderson & Wall, 1993; Cheng, 1998; Hughes, 1989; Watanabe, 2004). Some scholars also refer to this phenomenon as the direction of washback (Tsagari, 2009). It is noteworthy that this dimension of washback has close connection with the washback intentionality. In other words, depending on the context in which the washback effects are investigated, it is of utmost importance to identify whether the perceptions and practices of teachers and learners are of positive or negative value. In this regard, Watanabe (2004) has cautioned the washback researchers about labelling a practice as positive or negative as certain practices can be seen as positive by some stakeholders (for instance, principals) whereas similar practices can be seen as negative by others (for instance, teachers). Thus, Alderson (1992) suggested that in deciding the teaching and learning practices as either positive or negative, researchers need to consider 'who the evaluation is for'. In the case of the present study, the findings and discussion are meant for the MOE officials for policymaking in future and teachers to help them evaluate their teaching practices. Thus, the researchers primarily address the results of the current study in regard to what is intended by the MoE officials against what really occurs on the ground. The findings may also be useful for education systems that share similar policy concerns.

Methodology

Instrument

The questionnaire employed in this study was adapted from a validated teacher questionnaire from an earlier study looking into the washback effects of school-based oral assessment (Yu, 2010) in the context of Hong Kong. However, the original study administered this questionnaire to investigate the washback effects of school-based oral assessment (SBOA) on teaching and learning wherein an oral component was introduced into the curriculum. Therefore, the main focus of the questionnaire was on speaking skills. Constructs in the questionnaire representing Hong Kong context such as the demographic section and names of assessment components were replaced to suit the present (Malaysia) context. As the items in the adapted questionnaire focused on speaking skills (oral assessment) specifically, the researchers had to include additional items to ensure all four skills are equally represented. In addition, items were also adapted from another study in the context of Hong Kong (Cheng, 2005).

Also, the researchers altered the questionnaire from 6-point Likert scale to a 5-point Likert scale in order to fit the local context. Two experts with PhDs in the field of language testing examined the content validity of the adapted questionnaires, which is followed by making necessary changes in terms of wording. The questionnaire was then examined by two statisticians with M.A. degrees in statistics and eventually, it was piloted at a teacher training institution. Prior to the data collection of the main study, an internal consistency test of the teacher questionnaires revealed that their Cronbach's alpha values were at 0.96. The following table reveals the breakdown of the adapted survey instrument i.e., its sections, content, items and the sources they were adapted from:

Table 1

Breakdown of the Adapted Survey Instrument

Sections	Content	Items	Sources
Part I	*Demography*		
	Gender, teaching qualifications, years of teaching, etc.	1 - 13	Yu (2010) & Cheng (2005)
Part II	*Perceptions about School-based Assessment*		
	Overall impressions of PT3 and School Assessment components	14 - 36	
Part III	*Perceived Effects of School-based Assessment*		
	Perceived effects of PT3 assessment on teaching practices and students' learning		Yu (2010)
	Perceived effects of school assessment on teaching practices and students' learning	37 - 62	
Part IV	*Challenges of implementing school-based assessment in your school?*	63 - 72	Cheng (2005)
	Micro level challenges		
	Macro level challenges		
Total	**--**	**72**	**--**

Participants

The following table presents the information in relation to the survey administered in four different states in the northern region of Malaysia. The researcher administered the survey in

schools that volunteered to participate in the survey only. A total of (n) 143 survey instruments were sent out and its return rate was at 86.7%. In other words, only 124 research instruments were returned to the researcher.

Table 2
Return Rate of the Survey Instrument

State	Number of schools	Number of teachers
Penang	10	45
Perlis	4	18
Kedah	9	51
Perak	6	29
Total	29	143
Return rate	-	86.7% (124)

Ethical Issues

The researchers have to the best of their ability, sought to protect the rights and the identities of sampled participants. Before any data gathering was done, federal, state, and school ethical approvals were first obtained. Second, information such as the objective of the study, the amount of participation required, and the duration of the study were briefly explained to the teachers sampled for this study. Additionally, they were informed that their participation in the study was on a voluntary basis, and that they had the right to withdraw from it at any time, for any reason, and request removal of all or part of his/her data. Third, anonymity was given top priority with regard to the questionnaires completed by the participants of this study. All names, including those of the institutions and the people associated with them, were given as pseudonyms.

Results

The following table reveals the demographic information of the teacher participants who took part in the survey.

Table 3
Demographic Information

Demography		Frequency	Percentage (%)
Gender	Male	17	13.7
	Female	107	86.3
Age	20-29	15	12.1
	30-39	36	29.0
	40-49	48	38.7
	Above 50	25	20.2
Educational qualification	Bachelor's Degree	101	81.5
	Master's degree	16	12.9
	Other	7	5.6
I have been teaching English for	Less than 10 years	38	30.6
	More than 10 years	36	29.0
	Less than 20 years	7	5.6
	More than 20 years	43	34.7

My experience with SBA in English	Less than 5 years	96	77.4
	More than 5 years	7	5.6
	Less than 10 years	7	5.6
	More than 10 years	14	11.3
Forms currently teaching	Form 1	18	14.5
	Form 2	29	23.4
	Form 3	40	32.3
	Form 4	16	12.9
	Form 5	18	14.5
	Form 6	3	2.4
Type of school you are teaching	Daily school	121	97.6
	Boarding school	3	2.4
Did you opt to teach English as your preferred subject	Yes	95	76.6
	No	29	23.4

A cross-sectional survey was carried out in this study. It sampled (n = 124) English language teachers at the lower-secondary (form 3) level in the northern region i.e. Perlis, Perak, Kedah and Penang of Malaysia. By means of this survey (see Table 3), it was discovered that a striking 86.3% of the sampled respondents are female teachers and only 13.7% are males. In relation to their age, the highest number of respondents belonged to the category of 40-49 years old i.e. 38.7% whereas the other three categories which are 20-29 years old, 30-39 years old and above 50 years old had 12.1%, 29% and 20.2% respectively. With regards to ethnicity, 71% of the teachers are of Malay ethnic origin, 10.5% are Chinese, 16.1% are Indians and 2.4% are of other ethnic groups. Majority of the English language teachers hold bachelor's degrees as their academic qualification, about 12.9% of the respondents were master's degree holders and 5.6 % of the respondents hold other qualifications (Diploma in education, etc.). Majority of the sampled respondents had been teaching for more than 20 years i.e. 34.7%. Only 5.6% accounts for those serving less than 20 years. The other two categories, namely less than 10 years and more than 10 years had 30.6% and 29% respectively.

Since this study was carried out in the year of 2016 and considering the fact that SBA was introduced in the year of 2012 at the lower-secondary level (Form 3), 77.4% of the respondents chose to answer that their experience with SBA in English was less than 5 years. In a similar vein, 76.6% of them reported that they were part of the SBA committee in their respective schools. Among the six forms in secondary schools, majority of the respondents (32.3%) were teaching form 3 (Form 3) along with form 1 (14.5%), form 2 (23.4%), form 4 (12.9%), form 5 (14.5%) and form 6 (2.4%). In relation to the number of periods taught per week, the largest number of them (65.3%) were teaching between 22 and 27 hours. 76.6% of the sampled respondents chose to teach English as their preferred subject, whereas 23.4% of them were non-optionist English teachers.

As it was discussed in the literature review section of this paper, the present study only focused on the three significant features of the formative assessment i.e., teacher feedback or feedback, peer- and self-assessment. In this regard, the findings reported here are in relation to these three features of the formative assessment. In doing so, the teachers' perspectives with regards to

their practices are presented. They will be discussed and concluded in the following section. On reader-friendly grounds, teachers' teaching methods or practices were coded as TM (i.e., Teaching Methods) as it can be seen in the following:

TM1 *– I put more emphasis on giving my students feedback*
TM2 *– I use feedback on my students' SBA performance to improve my teaching*
TM3 *– I involve my students more in peer-assessment*
TM4 *– I involve my students more in self-assessment*

The Means procedure is employed to present the results of the present study as it is helpful for the description and analysis of scale variables in quantitative analysis. Within categories of one or more independent variables, the Means procedure computes subgroup means and related univariate statistics for dependent variables. The data was analysed using the Software Package for Social Sciences (SPSS, V21). The mean comparison test of this study comprises two independent variables (gender and experience in teaching) and one dependent variable (teaching method). The tables below illustrate the mean comparison of the teachers' gender and teaching experience mean score against the mean score of the teaching method:

Table 4
Gender vs Teaching Method (TM)

Gender		TM1	TM2	TM3	TM4
Male	Mean	3.59	3.76	3.65	3.53
	N	17	17	17	17
	Std. Deviation	.712	.664	.786	.800
Female	Mean	3.71	3.77	3.64	3.58
	N	107	107	107	107
	Std. Deviation	.740	.708	.827	.847
Total	Mean	3.69	3.77	3.65	3.57
	N	124	124	124	124
	Std. Deviation	.734	.700	.818	.838

It was found that female teachers were more inclined to offer feedback to their students compared to their male counterparts. A difference of 0.12 mean score can be observed between the two genders. With regards to the next item, it can be seen that both genders were almost at the same level in considering their students' performance in the school assessment component to improve their teaching. A subtle difference was observed though i.e., 0.01. A similar pattern was observed for the following item i.e., carrying out peer-assessment among the students. However, the female teachers were found to favour having their students engage in self-assessment. The difference in mean score was 0.05. Overall, both genders were equally positive about offering feedback to their students and using their students' performance as feedback for

their teaching methods. Also, a similar pattern was observed in administering peer- and self-assessment among their students.

Table 5

Experience in Teaching English vs Teaching Method (TM)

I have been teaching English for		TM1	TM2	TM3	TM4
less than 10 years	Mean	3.74	3.68	3.87	3.76
	N	38	38	38	38
	Std. Deviation	.685	.739	.875	.883
more than 10 years	Mean	3.69	3.83	3.61	3.61
	N	36	36	36	36
	Std. Deviation	.668	.609	.766	.766
less than 20 years	Mean	3.57	3.71	3.29	3.29
	N	7	7	7	7
	Std. Deviation	.535	.488	.488	.488
more than 20 years	Mean	3.67	3.79	3.53	3.42
	N	43	43	43	43
	Std. Deviation	.865	.773	.827	.879
Total	Mean	3.69	3.77	3.65	3.57
	N	124	124	124	124
	Std. Deviation	.734	.700	.818	.838

The findings of the study found that the teachers belonging to the category of having served less than twenty years had scored the lowest mean scores for the first item. On the other hand, teachers belonging to the category of having served less than ten years had scored the highest mean scores for the first item. As for the second item, it was found that teachers who had served for more than ten years scored the highest mean score. In comparison, the lowest mean score was by those who have served less than ten years. The others were in between these two categories. It was also found that the mean score differences were almost close to each category. As for the next item, the teachers who have served less than ten years scored the highest mean score i.e., 3.87. This difference in mean score between this category and the others was relatively bigger. For instance, the difference between this category and those who have served less than twenty years is 0.56. The other categories fell in between these two. Lastly, a similar pattern was observed for the last item as well.

Discussion

The findings of the present study revealed that female teachers at the lower-secondary level were generally more inclined to implement the three components of formative assessment, namely feedback, peer-, and self-evaluation among their students. Responses from male teachers were somewhat lower than those from female counterparts. The male teachers outperformed the female when it comes to adopting peer assessment. With reference to years of service as a teacher, those who had less than ten years had the highest mean scores for the

three categories; TM1, TM3, and TM4. Teachers who have served for fewer than twenty years followed a similar pattern with the lowest mean score.

The findings of this study is consistent with Alla Baksh et al. (2019), who also found that female instructors at the lower-secondary level were more concerned about the use of formative assessment at the lower-secondary level. Further, the findings of this study also echoed the findings in Shohamy (1996), in which teachers who were relatively new to the profession (for example, having worked for less than 10 years) follow evaluation policies more strictly than their more experienced colleagues. Prior to the implementation of formative assessment in the context of lower-secondary school in Malaysia, the practice of using test-derived information as feedback for teachers and students to adjust their teaching and learning activities was almost nonexistent as highlighted by Ong (2010). With the implementation of formative assessment, it is compulsory for teachers to administer school assessment by means of qualitative reporting of the students' language learning outcomes. As such, teachers are required to give their students constructive feedback and involve them in peer- and self-assessment. This may help serve the formative function of the school assessment more effectively. However, such a development may prompt concerns regarding the best way for teachers to explain to their students the discrepancy between the desired level of learning and the actual level of learning, as propagated by the concept of formative assessment.

As it can be seen in the findings of this study, similarities were found between both male and female teachers sampled in this study in terms of students' performance in school assessment component to evaluate their teaching methods. The researchers consider this a positive washback as the teachers (both males and females) consider the formative component (school assessment) in evaluating their own teaching methods. In this regard, it would be a welcome development if the educational authorities provide opportunities for teachers at the lower-secondary level to participate in training programs to enhance their understanding and implementation of formative assessment practices in the classroom. This may include providing training on how to develop and use formative assessment tools, as well as how to provide feedback to students that can help them improve their learning. Overall, similarities in attitudes towards feedback and assessment can be observed in the findings of this study. Despite some subtle differences, both male and female teachers were found to be equally positive about offering feedback to their students and using their students' performance as feedback for their teaching methods (Wall, 1996).

In terms of teaching experience, it was found that teachers with less than ten years of experience tended to employ a variety of teaching methods compared to those with more experience. It can be argued that those who are relatively new in the profession tend to explore various teaching methods compared to those experienced who may have attempted but feel the method/s is not effective in assessing the learning outcomes. The findings also point that the level of experience of a teacher may affect their effectiveness in certain areas, such as offering feedback to students. However, it is of utmost importance for the educational authorities to enhance the teachers' assessment literacy and support them in developing their skills over time and ensuring that they continue to grow and improve throughout their careers (Cheng, 2004).

Upon analyzing the results of the present study, it is evident that there are both commonalities and variances in the implementation of formative assessment in Malaysian lower-secondary education. Specifically, novice educators, as indicated by this study, demonstrate a greater inclination to experiment with diverse teaching methods as opposed to their more experienced counterparts. To optimize the evaluation of educational outcomes, it is recommended that educational authorities facilitate dialogues between both novice and experienced teachers.

Conclusion

This study has its limitations and it is imperative to highlight them in order to assist scholars that want to continue exploring the implementation of formative assessment raised in the present study accurately. Since questionnaire is used as the only instrument in the present study, methodological (interviews) and data triangulation (policymakers' and parents' perspectives) were not carried out to triangulate the results of the teachers' self-reported questionnaire responses. As such, future research may need to employ a mixed-methods approach in which both methodological and data triangulation can be carried out. Upon comprehensively reviewing the washback studies on the teaching methods employed by teachers (Alderson & Wall, 1993; Andrews, 2004; Cheng, 2004; Davison, 2008; Green 2007; Muñoz & Álvarez, 2010; Turner, 2009), it was discovered that most empirical investigations have reached the same conclusion, in which modifications to exams themselves may not always result in significant changes to teachers' instructional methods. Notably, if teachers are not given particular training or direction, this will only result in superficial adjustments to teaching approaches.

References

Alderson, J.C. (1992). Guidelines for the evaluation of language education. In Alderson, J.C. & Beretta, A. *Evaluating Second Language Education*. Cambridge University Press.

Alderson, J. C. (2004). The shape of things to come: Will it be the normal distribution? *Studies in Language Testing*, 18, 1-26.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.

Al Hinai, M. K., & Al Jardani, K. S. (2021). Washback in Language Testing: An Exploration with a Focus on a Specific EFL Context in Oman. *International Journal of English Linguistics*, 11(1), 138-147. https://doi.org/10.5539/ijel.v11n1p138

Alla Baksh, M. A., Mohd Sallehhudin A. 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), 1087-1104.

Alla Baksh, M. A., Abd Aziz, M. S., & Stapa, S. H. (2019). Examining the factors mediating the intended washback of the English language school-based assessment: Pre-service ESL teachers' accounts. *Pertanika Journal of Social Sciences & Humanities*, 27(1), 71-86. https://doi.org/10.47836/pjssh.v27i1.236

Alqahtani, N. (2022). English language usage and academic achievement among nursing students: A cross-sectional study. *SAGE Open Nursing*, 8, 2377968221109364. https://doi.org/10.1177/23779682211093643

Andrews, S. (2004). Washback and curriculum innovation. In D. Tsagari & J. I. Swain (Eds.), *Examining washback: The case of KPG exams in Greece* (pp. 59-72). Cambridge University Press.

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge University, School of Education.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.

Bailey, K. M. (1999). *Washback in language testing*. Educational Testing Service.

Berry, R. (2011). Assessment reforms around the world. In R. Berry & B. Adamson (Eds.), *Assessment Reform in Education* (Vol. 14, pp. 89-102). Springer.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.

Black, P., Harrison, C., Lee, C. & Marshall, B. (2003). *Assessment for Learning: Putting It into Practice*. McGraw-Hill International.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.

Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. Psychology Press.

Brindley, G. (2008). Educational reform and language testing. In *Encyclopedia of language and education* (pp. 2495-2508). Springer.

Butterfield, S., Williams, A. & Marr, A. (1999). Talking about assessment: mentor-student dialogues about pupil assessment in initial teacher training. *Assessment in Education: Principles, Policy & Practice*, 6(2), 225-246. https://doi.org/10.1080/09695949993044

Cheng, L. (1998). *The washback effect of public examination change on classroom teaching: an impact study of the 1996 Hong Kong Certificate of Education in English on the classroom teaching of English in Hong Kong secondary schools*. (Unpublished doctoral dissertation). The University of Hong Kong, Hong Kong.

Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In Savignon, S.J. (Ed.), *Interpreting Communicative Language Teaching: Contexts and Concerns in Teacher Education* (pp. 91-111). Yale University Press.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In Cheng, L., Watanabe, Y. & Curtis, A. (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 146-170). Lawrence Erlbaum Associates.

Cheng, L. (2005). *Changing Language Teaching through Language Testing: A Washback Study*. Cambridge University Press.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). Lawrence Erlbaum Associates.

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly*, 4(1), 37-68. https://doi.org/10.1080/15434300701352011

Dilova, N. G. (2021). Formative assessment of students' knowledge – as a means of improving the quality of education. *Scientific Reports of Bukhara State University*, 5(3), 144-155.

Fook, C. Y., Gurnam Kaur Sidhu, & Rizal Yunus. (2009). *School-based assessment enhancing knowledge and best practices*. University Publication Centre (UPENA), Universiti Teknologi MARA, Shah Alam, Malaysia.

Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95. https://doi.org/10.3102/00028312039001069

Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.

Hamp-Lyons, L. (1997). Washback, impact, and validity: Ethical concerns. *Language Testing*, 14(3), 295-303. https://doi.org/10.1177/026553229701400304

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Harris, L. R., Brown, G. T., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, 26(2), 107-133. https://doi.org/10.1007/s11092-013-9187-2

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7. https://doi.org/10.3102/0013189X020005002

Lee, M. N. (2006). Centralized decentralization in Malaysian education. In M. Bray, B. Adamson, & M. Mason (Eds.), *Educational Decentralization* (pp. 149-158). Springer.

Majid, Z. A., Samad, A. A., Muhamad, M., & Vethamani, M. E. (2011). The school-based oral English test: Similarities and differences in opinion between teachers and students. *The English Teacher*, 10, 113-128.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100. https://doi.org/10.1177/0265532208097337

Mok, M. M. C. (2012). Assessment reform in the Asia-Pacific region: The theory and practice of self-directed learning oriented assessment. In D. P. T. Tan, & J. W. Little (Eds.), *Self-directed Learning Oriented Assessments in the Asia-Pacific* (pp. 3-22). Springer.

Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49. https://doi.org/10.1177/0265532209348616

Ong, S. (2010). Assessment Profile of Malaysia: High-Stakes External Examinations Dominate. *Assessment in Education: Principles, Policy & Practice*, 17(1), 91-103. https://doi.org/10.1080/09695940903504938

Pearson, I. (1988). Tests as levers for change. In S. Chamberlain & R. Baumgardner (Eds.), *Language Testing: Principles and Applications* (pp. 98-107). Pergamon Press.

Popham, W. J. (1987). The merits of measurement-driven instruction. *The Phi Delta Kappan*, 68(9), 679-682.

Saville, N. (2009). Language assessment in the management of international migration: A framework for considering the issues. *Language Assessment Quarterly*, 6(1), 17-29. https://doi.org/10.1080/15434300902746021

Shepard, L. (1993). Interview on assessment issues with Lorrie Shepard. *Educational Researcher*, 20(2), 21–27. https://doi.org/10.3102/0013189X020002021

Shohamy, E. (1991). *Connecting Testing and Learning in the Classroom and on the Program Level*. In Northeast Conference on the Teaching of Foreign Languages.

Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. NFLC Occasional Paper. National Foreign Language Center.

Shohamy, E. (2001). Democratic Assessment as an Alternative. *Language Testing*, 18(4), 373-391.

Tan, H. M. (2009). *Changing the language of instruction for mathematics and science in Malaysia: The PPSMI policy and the washback effect of bilingual high-stakes secondary school exit exams* (Order No. NR66540). Available from ProQuest Dissertations & Theses Global. (808407479).

Tavares, N. J., & Hamp-Lyons, E. M. (2008). ESL teachers learning to assess interactively: The challenges and teacher development issues. In *Focusing on the Core*: *Justifying the Use of Language Assessments to Stakeholders*, 30th Annual Language Testing Research Colloquium, International Language Testing Association, Hangzhou, China, 23-28 June 2008.

Topping, K. J., & Ehly, S. W. (Eds.). (1998). *Peer-Assisted Learning*. Lawrence Erlbaum.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. *In Optimising New Modes of Assessment: In search of Qualities and Standards* (pp. 55-87). Springer.

Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. *Center for Research in Language Education, CRILE Working Papers*, 58, 1-23.

Turner, C. E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, 5(1), 103-123.

Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13(3), 334-354.

Watanabe, Y. (2004). Teacher factors mediating washback. In *Washback in Language Testing: Research Contexts and Methods* (pp. 129-146). Lawrence Erlbaum Associates.

Widen, M. F., O'shea, T., & Pye, I. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4), 428-444.

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228-260.

Yu, Y. (2010). *The Washback Effects of School-based Assessment on Teaching and Learning: A Case Study*. (Order No. NR66540). Available from ProQuest Dissertations & Theses Global. (808407479).

Acknowledgments

Not applicable.

Funding

Not applicable.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which grants permission to use, share, adapt, distribute and reproduce in any medium or format provided that proper credit is given to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if any changes were made.