Pertanika J. Soc. Sci. & Hum. 27 (T2): 47 - 62 (2019)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Assessment for Learning: Espoused and Enacted Practices of Malaysian Teachers

Renuka V. Sathasivam[1], Moses Samuel[2], Norjoharuddeen Mohd Nor[1], Meng Yew Tee[3] and Kwan Eu Leong[1]*

[1]Department of Mathematics and Science Education, Faculty of Education, University Malaya, Kuala Lumpur 50603, Malaysia
[2] School of Education, Taylor's University, Selangor 47500, Malaysia
[3]Department of Curriculum and Instructional Technology, Faculty of Education, University Malaya, Kuala Lumpur 50603, Malaysia

ABSTRACT

High stakes assessment practices have been the most common summative assessments in schools. To address the shortcomings of the summative assessments, Assessment for Learning (AfL) was introduced to provide learners more active feedback and autonomy in their learning progression. In Malaysia, school-based assessments were introduced to reduce the dependency on summative assessments. This paper examines the espoused and the levels of enacted AfL practices of Malaysian Year 7 teachers. A survey and video research design was employed in this study where data were collected from a sample of teachers who taught Mathematics, Science, English Language, and Malay Language. To elicit teachers' espoused AfL practices, a questionnaire was used while video recordings were utilized as a form of classroom observation to elicit teachers' enacted AfL practices. The results indicated that there seemed to be agreement among teachers' espoused AfL practices with AfL strategies for three dimensions which were Sharing Learning Target, Engineering Good Classroom Discussions, and Peer Assessment. The enacted AfL practices of these teachers were predominantly at the lowest levels for all dimensions of AfL practices. This study was not able to establish any relationship between the espoused and enacted AfL

ARTICLE INFO

Article history:
Received: 8 August 2017
Accepted: 10 April 2018
Published: 15 May 2019

E-mail addresses:
renukasivam@um.edu.my (Renuka V. Sathasivam)
Moses.Samuel@taylors.edu.my (Moses Samuel)
norjo@um.edu.my (Norjoharuddeen Mohd Nor)
mytee@um.edu.my (Meng Yew Tee)
rkleong@um.edu.my (Kwan Eu Leong)
* Corresponding author

ISSN: 0128-7702
e-ISSN: 2231-8534

© Universiti Putra Malaysia Press

practices. It is suggested that both, espoused and enacted AfL practices were be brought to light as foundations for continuous improvement. It is hoped that these results would inform and influence policy makers about the future direction and decision making in improving the quality of school-based assessment implementation.

Keywords: Assessment for learning, enacted practices, espoused practices, Malaysia, school-based assessment

INTRODUCTION

Within the Asian region, reforms on assessment practices have been undertaken to address the shortcomings of high-stakes summative assessments. In order to achieve this, countries such as India, Hong Kong, Vietnam, Singapore, and Malaysia have begun to move away from high-stakes summative assessments towards formative assessments (Chan, 2015; Koh & Luke, 2009; Ministry of Education [MOE], 2012; National Council of Educational Research and Training [NCERT], 2005; Thanh-Pham & Renshaw, 2014). One form of formative assessment that is being promoted is Assessment for Learning (AfL) which is conceptualized as a "process of seeking and interpreting evidence [of learning] for use by learners and their teachers to decide where the learners are in their learning, where they need to go, and how best to get there" (Assessment Reform Group [ARG], 2002). AfL aspires to give students autonomy in their learning progression by employing processes such as sharing learning targets and criteria of success with students; providing constructive feedback; and promoting students' active involvement in assessment (Swaffield, 2011; Thomas et al., 2011).

This reform in assessment requires a shift, not only in teachers' knowledge of AfL, but also in their classroom practices. The relationship between teachers' AfL knowledge which can be called their 'espoused AfL practices' and their enacted AfL practices need to be investigated. Reforms in assessment have the potential to destabilize teachers' espoused-enacted practices' alignment; the congruence between them is especially crucial in ensuring the success of these reforms (Dixon & Haigh, 2009; Richardson et al., 1991). It is the incongruences, if any, between the espoused and enacted AfL practices that have to be illuminated and acted on so that sufficient support can be provided to help teachers align their espoused and enacted practices.

Literature Review and Theoretical Foundation

AfL principles are often guided by Sadler's (1989) framework that states teachers need to establish where students are going in their learning, where they are right now and how best to achieve their learning goals, and to involve students in each and every step of the way. The AfL principles have to remain relatively stable with the goal to involve students in the assessment process (Andrade & Brown, 2016; Black & Wiliam, 1998). Research on AfL recommends that teachers

share explicit and understandable description of the learning targets with their students (Black & Wiliam, 2009; Chappius, 2015). Teachers are advised to use exemplars of good and weak work and to explain the differences between them. Teachers should also provide regular descriptive feedback that focuses on the strengths and weaknesses of student work and the kind of learning that needs to be addressed (Hattie & Timperley, 2007). By modelling these practices (i.e., sharing learning targets and providing descriptive feedback), teachers are inviting students into the assessment process and making it transparent to them. This, in turn, gives students the ability to internalize the assessment criteria and the knowhow to use the information to do self- or peer-assessment and then communicate their ideas or feedback to their peers. All this can be done successfully only if teachers create relevant and meaningful tasks for their students (Heritage, 2010). However, research seems to suggest that how these strategies are implemented in the classrooms and how they complement each other is complex (DeLuca et al., 2012; Klenowski, 2009; Wiliam & Thompson, 2007).

Research on teachers' espoused practices, beliefs, and knowledge about AfL have been premised on the assumption that these are predictors of what actually happens in classrooms (Atjonen, 2014; Irving et al., 2011; Jonsson et al., 2015; Remesal, 2007; Sach, 2012). However, due to the lack of concrete interpretation of exactly what is AfL and how it might work in a real-life setting (Dunn & Mulvenon, 2009; Swaffield, 2011), teachers' espoused AfL practices might not be what they actually enact in their classrooms. Teachers may believe that what they are practicing are indeed the AfL strategies. As such, teachers may espouse that they share learning targets with their students by providing exemplars; however, if they do not explicitly link the exemplars with the assessment criteria, the assessment criteria will still be inaccessible to the students (Booth et al., 2014; Sadler, 2010). Furthermore, enactment of AfL strategies may be inferior as teachers tend to engage superficially when their beliefs are not aligned with the underlying principles of AfL (Dixon et al., 2011; Marshall & Drummond, 2006). Marshall and Drummond (2006) investigated how teachers translated AfL strategies into practice and found that although the teachers were able to espouse AfL principles, not all teachers were able to enact the practices in the 'spirit of AfL'.

The (mis)alignment between teachers' espoused and enacted AfL practices needs to be addressed because in their seminal Theory-of-Action, Argyris and Schon (1974) argued that the effectiveness of human actions resulted from congruence between the espoused theory (what they say) and their theory-in-use (what they do). In the case of teachers, their enacted practices provide an indication of their theory-in-use, i.e., the tacit knowledge that drives their actions (Smith, 2001). By uncovering teachers' enacted practices, the implicit theories will be made explicit (Marland, 1995). This would make teachers to be more aware of the theoretical underpinning

of their actions and provide the foundations for continuous improvement (Dixon et al., 2011; Schön, 1983). As such, studies that examine what teachers say or believe (espouse) about assessment runs the risk of revealing only 'half the story' because it cannot be assumed that what teachers espouse is what they are practicing in their classrooms (Kane et al., 2002; Willis, 1993). As such research that uses self-report questionnaires to elicit teachers' espoused practices or beliefs should be confirmed with classroom observations (Brown et al., 2015; Lucero-Mareydt et al., 2013). It is, therefore, crucial to investigate both teachers' espoused and enacted practices.

School-based Assessment in Malaysia

In recent years, efforts have been made to move away from over-dependence on these summative high-stakes public examinations toward a formative school-based assessment. For example, Malaysia had introduced a school-based assessment initiative known as *Pentaksiran Berasaskan Sekolah* (School-based Assessment, hereafter referred to as SBA) which was implemented in stages, beginning with Year 1 in 2011 and Year 7, the first year of secondary school, from 2012. The SBA is a systematic assessment procedure conducted by subject teachers during the teaching and learning process following guidelines from the Malaysian Examination Syndicate (*Lembaga Peperiksaan Malaysia*, in Malay) in line with the National Education Assessment System (NEAS). The main objectives of the NEAS are to formulate and strategize on how to reduce the focus on public examinations, to improve students' learning, to create holistic assessment, to develop better human capital, and to strengthen school-based assessment (Lembaga Peperiksaan Malaysia, 2011).

This school-based assessment initiative aims to give more autonomy for teachers to design forms of assessment tailored for their particular teaching-learning needs; to use feedback from assessment to diagnose learning problems; to recalibrate teaching strategies; and to report the outcomes of assessment to other stakeholders (Kamal & Rahman, 2006; Md-Ali et al., 2015). The empowerment accorded to teachers in this school-based assessment system poses a number of challenges (Norzila, 2013). In Malaysia, there were strong murmurs among the rank and file teachers about the difficulties of implementing school-based assessment. Many teachers struggled with the administrative requirements imposed by the Ministry of Education, but other issues raised were actually quite commonly experienced by teachers in general. Some major obstacles include class size, curriculum requirements, lack of resources in developing the formative assessment process, as well as the possible tensions between school-based formative assessments and high visibility summative examinations (Brown et al., 2009; OECD, 2005). In the first four years of its implementation little is yet known about what actually takes place in the classrooms with regards to SBA. A key issue in the Malaysian context is whether these assessment practices under SBA are consistent with AfL practices.

Purpose of Research

The purpose of the research was to test the hypotheses that there is a relationship between Malaysian teachers' espoused and enacted AfL practices. This paper examines the espoused AfL practices and the levels of enacted AfL practices of Malaysian Year 7 teachers and intends to answer the following research questions: (1) To what extent are the espoused AfL practices of teachers aligned/ consistent with enacted AfL practices? (2) What are the levels of the enacted AfL practices of these teachers? (3) Is there a significant relationship between teachers' espoused AfL practices and the levels of enacted AfL practices in terms of (a) Sharing Learning Targets, (b) Engineering Good Classroom Discussions, (c) Descriptive Feedback and (d) Peer Assessment?

METHODOLOGY

This paper reports part of a larger on-going national-scale research investigation known as 'Inquiry into Malaysian Classroom Educational Practices' or IMCEP. It was undertaken to describe a bird's eye view of classroom educational practices in Malaysia (Tee et al., 2016). A survey and video research design were employed in this study where data was collected from a sample and results of the analysis were generalized to the population. The population of this study was Year 7 teachers of Mathematics, Science, English Language, and Malay Language in Malaysia. These are the core subjects taught in the Malaysian educational system. There were two key reasons why Year 7 was chosen. Firstly, Year 7 was the transition year from primary to secondary school – allowing researchers the insights into students' classroom experience just after completing primary school and at the beginning of secondary school. Year 7 would set the tone for these students as they progress towards following years in secondary school. Secondly, at the point of data collection in 2014, these teachers would have had a minimum of two years of experience in implementing SBA and thus would have assumed some measure of stability in their classroom assessment practices. The sampling of these teachers began with a computer-aided random sampling of approximately 2000 public secondary schools in Malaysia which enrol about 88% of the Malaysian secondary school student population (MOE, 2012). From this list of randomly sampled schools, the researchers requested permission from the necessary local authorities and the voluntary participation teachers in the respective schools. In total, 153 teachers from the 24 schools participated in this study. However, only 121 teachers had the complete set of data comprising the survey and video recordings of three lessons. On the average there were about 30 teachers for each subject. The data from these teachers was analysed and is reported in this paper.

Survey and Video Data

To elicit teachers' espoused AfL practices, Question 28 – with 10 sub items – from the IMCEP questionnaire was used. The researchers refer to Question 28 as *Inquiry into Malaysian Classroom Educational*

Practices-Assessment for Learning or IMCEP-AfL. The questions in IMCEP-AfL were to gain an insight into teachers' espoused AfL practices. Teachers responded to each sub item using a four-point Likert scale from 1= "strongly disagree" to 4= "strongly agree".

For construct validity, an exploratory factor analysis was conducted on the 10 items from IMCEP-AfL. In this analysis, principal component extraction was used to obtain the factor structure of the items. Using the principal component analysis and variamax rotation produced four factors with eigen values greater than one. Items with loadings greater than 0.5 were retained in the four factors. The four factors contributed 51% of the variance in the respondents' score. The factor pattern showed that (a) Factor 1 was measured by items 28a and 28c and was labelled Sharing Learning Targets; (b) Factor 2 was measured by items 28e, 28j and 28k and was called Engineering Good Classroom Discussions; (c) Factor 3 was represented by items 28h and 28p and was named Descriptive Feedback; and (d) Factor 4 was represented by items 28f, 28g and 28n and was labelled Peer Assessment.

The reliabilities of these items were calculated using the Cronbach's alpha value. Table 1 displays the constructs, number of items and the Cronbach alpha coefficients. The internal reliability values were in the range of 0.52 to 0.63 showing sufficient internal consistency for IMCEP-AfL.

Table 1

Dimensions of constructs of IMCEP-AfL with sample items

Construct	Item number	Cronbach alpha reliabilities	Sample items
Sharing Learning Targets	2	0.53	28a. I let my students know what they will be learning (assessed) for the day.
Engineering Good Classroom Discussions	3	0.63	28e. I ask questions to my students to gain information about their understanding.
Descriptive Feedback	2	0.52	28h. I usually do not tell my students the right answers.
Peer Assessment	3	0.52	28g. I provide opportunities in my class for my students to know what their peers are thinking or their understanding of a topic.

Video data was collected to gain an insight into teachers' enacted AfL practices. For each teacher, three lessons were video recorded over the span of a week. Each lesson lasted one to three periods (approximately 40-120 min). Two video cameras (one stationary, placed in front of the classroom, and the other primarily followed the teacher)

and one audio recorder (attached to the teacher) were utilized to record each lesson. There were usually two research assistants present during the recordings.

Data Analysis Procedure

The teachers' espoused AfL practices were measured from the responses to the items in ICMEP-AfL. Using descriptive quantitative analysis, percentages of teachers' responses to the statements were computed. The teachers' enacted AfL practices were measured from the video recordings which were then analysed using *a priori* coding framework called Malaysian Teachers' Assessment Practices Instrument (or MTAPI) which was developed for the IMCEP study. The analysis of the video data using MTAPI was focused on how the observed practice most resembled established good practices, as defined by AfL literature (Black & Wiliam, 2006, 2009; Wiliam, 2006, 2010; Wiliam & Thompson, 2007). The dimensions in MTAPI included (i) Sharing learning target (ii) Engineering good classroom discussions, (iii) Descriptive feedback, and (iv) Peer assessment. The levels of each teacher's practice were coded as 'Unsatisfactory', 'Basic', 'Proficient', and 'Distinguished'. For example, a teacher who did not write or say the learning outcomes for the lesson will be coded as 'Unsatisfactory' for Sharing Learning Targets dimension, which has the descriptor: *The teacher does not convey or makes no effort to convey the learning targets to students.* Table 2 shows the Sharing Learning Targets dimension, the descriptors and examples of teachers' practices. The assessment practices were coded by two coders by mutual consent. For details of how this process was developed, please refer to Tee et al. (2016).

Table 2

Example of MTAPI for the dimension Sharing Learning Targets

	Dimension	Unsatisfactory	Basic	Proficient	Distinguished
A.	**Sharing Learning Targets**	The teacher **does not convey** or **makes no effort** to convey the **learning targets** to students.	The teacher **conveys** the **learning targets** in: ▪ Monologue fashion or purely procedural manner ▪ Limited participation by students	The teacher **conveys and explains** the **learning targets** in: ▪ Student-friendly language ▪ With participation of students ▪ Encouraging ways	The teacher **conveys and explains** the **learning targets** in: ▪ Students-friendly language ▪ This is shared with all students. ▪ Encouraging ways ▪ Students and teachers establish individualized learning goals

Table 2 *(Continued)*

Examples			
(a) Teacher does not write/say/explain the learning objectives/ outcomes to the class	(a) Teacher writes/says the learning objectives/ outcomes to the class but does not explain the learning objective/ outcomes to the class.	(a) Teacher writes/says and explains the learning objectives and outcomes to the class in a friendly manner	(a) Teacher discuss and set the learning goals/outcomes with all students in a friendly manner (comes with examples)
(a) Teacher does not explain what are the elements (b) that will be assessed	(a) Teacher does not explain what are the elements that will be assessed	(a) Teacher encourages students and plan strategies to help students to achieve learning objectives/outcomes	(a) Teacher encourages students and plan strategies to help students to achieve learning objectives/outcomes
		(a) Teacher sets the learning objectives/ outcomes that matches all students' learning abilities	(a) Teacher asks students to define/set their own learning targets
		(a) Teacher discusses with students in determining the elements to be assessed based on the learning targets globally	(a) Teacher asks students to determine the elements to be assessed.

Validity and reliability procedures for video data coding were carried out at multiple levels. Firstly, training was carried out so that coders could "get on the same page" in the way MTAPI was utilized. Secondly, a paired-coding system was installed. Two coders would watch the same video, and then coded the video by consent. Thirdly, a quantitative post hoc approach was used to measure reliability score. The coding by experts and the coders were found to be significantly correlated at $p < 0.0001$, based on the Single Measures Intraclass Correlation Coefficient (0.631).

FINDINGS

This section presents the findings, based on the analysis, in order to answer the three research questions. For the first research question, the results of the descriptive analysis of the espoused AfL practices of teachers are shown in Table 3.

The findings tabulated in Table 3 show that more than 80% of the sample at least agreed to the statements for the dimensions of Sharing Learning Targets, Engineering Good Classroom Discussions, and Peer Assessment. For example, for the dimension of Sharing Learning Target 23.1% and 66.1% strongly agreed and agreed, respectively, that they would let their students know what they (the students) would be learning for that lesson. Another example would be for the dimension of Engineering Good Classroom Discussions, where 22.3% and 71.1% of the teachers strongly agreed and agreed, respectively, that they held discussions to elicit useful information on their student learning. Whereas, less than 10% at least agreed to the statements for the dimension

Table 3
Descriptive analysis of the espoused AfL practices of teachers

Criteria	Item (IMCEP-AfL)	Espoused											
		Strongly Agree			Agree			Disagree			Strongly Disagree		
		F	(%)	95% CI	F	(%)	95% CI	F	(%)	95% CI	F	(%)	95% CI
Sharing Learning Targets	28a. I let my students know what they will be learning (assessed) for the day.	28	23.1	[15.6,30.6]	80	66.1	[57.7, 74.5]	10	8.3	[3.4, 13.2]	3	2.5	[0.0, 5.3]
	28c. I show examples and explain to my students what standard they should be aiming for in order to reach their learning goals.	28	23.1	[15.6,30.6]	89	73.6	[65.8, 81.5]	4	3.3	[0.1, 6.5]	0	0.0	[0.0, 0.0]
Engineering Good Classroom Discussions	28e. I ask questions to my students to gain information about their understanding.	43	35.5	[27.0, 44.0]	74	61.2	[52.5,69.9]	3	2.5	[0.0, 5.3]	1	0.8	[0.0, 2.4]
	28j. I hold discussions in my lessons because it is an important source of evidence of student learning.	27	22.3	[14.9, 29.7]	86	71.1	[63.0,79.2]	8	6.6	[2.2,11.0]	0	0.0	[0.0, 0.0]
	28k. Even if my students have given me the right answer, I will still ask them to explain their answers.	22	18.2	[11.3,25.1]	91	75.2	[67.5, 82.9]	8	6.6	[2.2,11.0]	0	0.0	[0.0, 0.0]
Descriptive Feedback	28h. I usually do not tell my students the right answers.	2	1.7	[0.0,4.0]	7	5.8	[1.6, 10.0]	72	59.5	[50.8, 68.3]	40	33.1	[24.7, 41.5]
	28p. I don't feel at ease when I have to constantly provide grades for my students' learning.	0	0.0	[0.0, 0.0]	9	7.4	[2.7, 12.1]	89	73.6	[65.8, 81.5]	23	19.0	[12.0, 26.0]
Peer Assessment	28f. I think my students are capable of assessing their friends.	10	8.3	[3.4, 13.2]	74	61.2	[52.5 ,69.9]	36	29.8	[21.7,38.0]	1	0.8	[0.0,2.4]
	28g. I provide opportunities in my class for my students to know what their peers are thinking or their understanding of a topic.	17	14.0	[7.8, 20.2]	77	63.6	[55.0, 72.2]	25	20.7	[13.5, 27.9]	2	1.7	[0.0, 4.0]
	28n. I think it is right for students to comment about the quality of their peers' work.	12	9.9	[4.6, 15.2]	55	45.5	[36.6, 54.4]	48	39.7	[31.0, 48.4]	6	5.0	[1.1, 8.9]

of Descriptive Feedback. For instance, only 1.7% and 5.8% of the teachers strongly agreed and agreed, respectively, that they should not tell their students the right answers and it can be interpreted that a majority of these teachers felt that they should let their students know the right answers. Therefore, the results show that the teachers' espoused practices are aligned with three out of four dimensions of AfL practices.

For the second research question, the results of the descriptive analysis of the enacted AfL practices of teachers are shown in Table 4.

Referring to Table 4, the results suggest that more than 90% of the participating teachers' AfL practices for all the four AfL dimensions were not at the 'Distinguished' or 'Proficient' level. For example, for the dimension of Sharing Learning Targets, 81.8 % and 18.2% of the teachers were categorized as 'Unsatisfactory' and 'Basic', respectively. For the dimension of Engineering Good Classroom Discussions, 75.2% and 24.0 % of the teachers were at the 'Unsatisfactory' and 'Basic' level, respectively. For the dimension of Descriptive Feedback, 55.4% and 41.3% of the teachers were labelled as 'Unsatisfactory' and 'Basic' level. Similarly, for the dimension of Peer Assessment 92.6% and 7.4% of the teachers were found to be at the 'Unsatisfactory' and 'Basic' level. Therefore, the results show that the teachers' enacted AfL practices were predominantly at the lowest level for three dimensions, namely Sharing Learning Targets, Engineering Good Classroom Discussions, and Peer Assessment.

Table 4
Descriptive analysis of the enacted AfL practices of teachers

Criteria	Enacted											
	Distinguished			Proficient			Basic			Unsatisfactory		
	F	(%)	95% CI	F	(%)	95% CI	F	(%)	95% CI	F	(%)	95% CI
Sharing Learning Targets	0	0.0	[0.0,0.0]	0	0.0	[0.0, 0.0]	22	18.2	[11.3, 25.1]	99	81.8	[74.9, 88.7]
Engineering Good Classroom Discussions	0	0.0	[0.0,0.0]	1	0.8	[0.0,2.4]	29	24.0	[16.4, 31.6]	91	75.2	[67.5, 82.9]
Descriptive Feedback	0	0.0	[0.0,0.0]	4	3.3	[0.1, 6.5]	50	41.3	[32.5, 50.1]	67	55.4	[46.5, 64.3]
Peer Assessment	0	0.0	[0.0,0.0]	0	0.0	[0.0, 0.0]	9	7.4	[2.7, 12.1]	112	92.6	[87.9, 97.3]

To answer the third research question, the results are presented in Table 5.

Referring to Table 5, at 5% significant level, the result of the Spearman Correlation Coefficient test showed that there is statistically significant correlation between the espoused and the enacted AfL practices by teachers, r (121) = 0.188, p = 0.039 for the dimension of Engineering Good Classroom Discussions. This correlation has a small positive effect size (Cohen, 1988). Thus, it can be concluded that there is a relationship between espoused practices by the teachers and their enacted practices in terms of Engineering Good Classroom Discussions. However, for the other dimensions of AfL, there are no statistically significant correlations between teachers espoused and enacted AfL practices in terms of Sharing Learning Targets, r (121) = -0.003, p = 0.976, Descriptive Feedback , r (121) = 0.106, p = 0.248 and Peer Assessment, r (121) = 0.172, p = 0 .059. Therefore, there is not enough evidence from the data to conclude that there is a relationship between teachers' espoused and enacted AfL practices in terms of Sharing Learning Targets, Descriptive Feedback, and Peer Assessment.

Table 5

Relationship between espoused and enacted AfL practices

	Espoused Sharing Learning Targets	Espoused Engineering Good Classroom Discussions	Espoused Descriptive Feedback	Espoused Peer Assessment
Enacted Sharing Learning Targets	-0.003	-	-	-
Enacted		0.188*	-	-
Engineering Good Classroom Discussions	-			
Enacted	-	-	0.106	-
Descriptive Feedback				
Enacted	-	-	-	0.172
Peer Assessment				

Note. N=121; Spearman Rho correlations are significant at $*p \leq 0.05$

DISCUSSION AND CONCLUSION

The research began by hypothesizing that there is a relationship between Malaysian teachers' espoused and enacted AfL practices. The findings of the study, however, showed that the hypothesis is not true for Malaysian teachers for three out of the four AfL dimensions, namely, Sharing Learning Targets, Descriptive Feedback, and Peer Assessment. Although the espoused-enacted relationship in the Engineering Good Classroom Discussions dimension was statistically significant, the effect size was small. In essence, teachers' espoused AfL practices are not related to their enacted practices in all four dimensions. Alignment and coherence between teachers' espoused and enacted are significant because it can have a positive influence on the quality of student learning outcomes (Sandvoll, 2014).

There could be various explanations for the misalignment of the espoused-enacted practices. Firstly, any reform in assessment has the potential to destabilize teachers' espoused-enacted practices (Dixon & Haigh, 2009; Richardson, et al., 1991). Since it was still early in the reform process (in the third year of implementation), the findings of this study seem to suggest that teachers were able to espouse practices that were consistent with AfL strategies, but these were not aligned to their enacted practices. For instance, for the dimension Sharing Learning Targets, teachers espoused that they would let their students know what they will be learning (assessed) for the day. These teachers also espoused that they would show examples and explain to their students what standards they (the students) should be aiming for in order to reach their learning goals. However, video data showed that teachers often proceeded with the lesson without discussing the learning outcomes with the students. In other words, the teachers were able to espouse practices that are aligned with AfL principles but were unable to enact them (DeLuca et al., 2012; Marshall & Drummond, 2006). As teachers are familiarizing themselves with SBA, it may be easier for the teachers to espouse their practices but may require more time to be fully comfortable with their actual AfL practices or develop sufficient skills to put the new knowledge into practice. This is consistent with other studies that have also found that although teachers appear to be committed to AfL concepts and strategies, they still struggle to enact the practices as they try to put their AfL knowledge into practice (Dixon et al., 2011; Irving et al., 2011; Klenowski, 2009; Marshall & Drummond, 2006; Willis, 2011).

Secondly, while teachers may espouse what they believe, they may not however have the knowledge and competency to enact these practices. In this study, teachers' enacted practices for the Dimension Peer Assessment was the highest in the 'Unsatisfactory' level with a value of 92.6%. The video data showed teachers do not encourage students to share and discuss their work or even evaluate their peers' work. The implementation of peer assessment in the classroom is a complex challenge because teachers are reluctant to relinquish complete control over the feedback and assessment process (Kember, 2009;

Nicol et al., 2014). It is also possible that teachers may enact traditional assessment practices, despite their beliefs that AfL is a good assessment system, as they perceive these are easier to implement as suggested by Giddens's structuration theory (Giddens, 1984). To them, the assessment practices prior to the implementation of the SBA demand less disruptions and friction with real-life factors such as resources, curricula, student readiness and/or time.

As the findings indicate that there is lack of alignment between teachers' espoused and enacted practices, the challenge lies in how we can change teachers' enactments: How can we change teachers' practices so that they are reflective of AfL principles? Literature suggests that one cannot assume espoused practices are indicative of teachers' beliefs. Teachers' beliefs are often more difficult to change. Thus, in introducing any reforms due care must be taken to ensure teachers' belief systems are addressed. Hunzicker (2004) suggested that "the change process involves a slow progression through stages that lead to eventual readiness for change". While educational innovations in Malaysia are generally top-down driven as it is a highly centralized system, venues must be explored to involve teachers more directly in the change process. As far as AfL is concerned, the teachers need to be inducted into the process of Sharing Learning Targets, Engineering Good Classroom Discussions, Descriptive Feedback, and Peer Assessment. As Argyris and Schon (1974) argued, the effectiveness of teachers' practices resulted from congruency between their espoused and enacted practices i.e. between what they said and what they did. The way forward may thus involve bringing both - the espoused and the enacted-practices to light, in order to help teachers to become more aware of the theoretical underpinnings of their practices and provide the foundations for continuous improvement.

ACKNOWLEDGEMENT

This work was funded in part by the University of Malaya Research Grant (UMRG) RP004-13SBS, the Equitable Society Research Cluster, and the University of Malaya Rakan Penyelidikan Grant CG0352013.

REFERENCES

Andrade, H. L., & Brown, G. T. L. (2016). Student self-assessment in the classroom. In G. T. L. Brown & L. R. Harris (Eds.), *The handbook of human and social conditions in assessment* (pp. 319-334). New York, USA: Routledge.

Argyris, C., & Schon, D. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, USA: Jossey-Bass.

Assessment Reform Group. (2002). *Assessment for learning*. Retrieved April 16, 2015, from http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf

Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal*, *25*(2), 238-259.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81-100). London, England: Sage.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment of Evaluation and Accountability*, *21*, 5-31.

Booth, B., Hill, M. F., & Dixon, H. (2014). The assessment-capable teacher: Are we all on the same page? *Assessment Matters*, *6*, 137-157.

Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*, *71*, 50-64.

Chan, C. K. Y. (2015). Use of animation in engaging teachers and students in assessment in Hong Kong higher education. *Innovations in Education and Teaching International*, *52*(5), 474-484.

Chappius, J. (2015). *Seven strategies for assessment for learning* (2nd ed.) New York, USA: Pearson Education.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, USA: Lawrence Erlbaum Associates.

DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters*, *4*, 5-29.

Dixon, H., & Haigh, M. (2009). Changing Mathematics teachers' conceptions of assessment and feedback. *Teacher Development: An International Journal of Teachers' Professional Development*, *13*(2), 173-86.

Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 365-379.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment Research & Evaluation*, *14*(7). Retrieved December 22, 2014, from http://pareonline.net/getvn.asp?v=14&n=7

Giddens, A. (1984). *The constitution of society. Outline of the theory of saturation*. Cambridge, England: Polity Press.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, USA: Corwin Press

Hunzicker, J. (2004). The beliefs-behavior connection: Leading teachers toward change. *Principal*, *84*, 44-46.

Irving, S. E., Harris, L. R., & Peterson, E. R. (2011). 'One assessment doesn't serve all the purposes' or does it? New Zealand teachers describe assessment and feedback. *Asia Pacific Education Review*, *12*, 413-426.

Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 104-121.

Kane, R., Sandretto, S., & Heath, C. (2002).Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, *72*(2), 177-228.

Kember, D. (2009). Promoting student-centred forms of learning across an entire university. *Higher Education*, *58*, 1-13.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 263-268.

Koh, K. & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 291-318.

Lembaga Peperiksaan Malaysia. (2011). *Panduan dan peraturan PBS* [PBS guidelines and rules]. Putrajaya, Malaysia: Ministry of Education.

Lucero-Mareydt, M., Valcke, M., & Schellens, T. (2013). Teachers' beliefs and self-reported use of inquiry in science education in public primary schools. *International Journal of Science Education*, *35*(8), 1407-1423.

Marland, P. (1995). Implicit theories of teaching. In L. W. Anderson (Ed.), *International encyclopaedia of teaching and teacher education* (2nd ed., pp. 131-136). New York, USA: Pergamon.

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, *21*(2), 133-149.

Md-Ali, R., Veloo, A., & Krishnasamy, H. N. (2015). Implementation of school-based assessment: The experienced teachers' thoughts. *Australian Journal of Basic and Applied Sciences*, *9*(18), 72-78.

Ministry of Education. (2012). *Malaysia education blueprint 2013-2025*. Putrajaya, Malaysia: Ministry of Education, Government of Malaysia.

National Council of Educational Research and Training. (2005). *National curriculum framework 2005*. New Delhi, India: NCERT.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education*, *39*(1), 102-122.

Norzila, M. Y. (2013). School-based assessment: Transformation in educational assessment in Malaysia. Cambridge Horizons School-based Assessment: Prospects and Realities in Asian Contexts. Retrieved February 10, 2015, from http://www.cambridge assessment .org.uk/images/139719-sba-seminar-papers.pdf

Organisation for Economic Co-operation and Development. (2005).*Formative assessment: Improving learning in secondary classrooms*. Retrieved December 22, 2014, fromhttp://www.oecd.org/edu/ceri/35661078.pdf

Kamal, R. Z., & Rahman, S. A. (2006, May 18). Sistem peperiksaan diubah: Tumpu penilaian seimbang kemahiran pelajar selain akademik [Changes in examination system: Focus on a balanced evaluation of students' skills other than academic]. *Utusan Malaysia*, p.7.

Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *The Curriculum Journal*, *18*(1), 27-38.

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, *28*(3), 559-586.

Sach, E. (2012). Teachers and testing: An investigation into teachers' perceptions of formative assessment. *Educational Studies*, *38*(3), 261-276.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, *35*(5), 535-550.

Sandvoll, R. (2014). When intentions meet reality: Consonance and dissonance in teacher approaches to peer assessment. *Canadian Journal of Higher Education*, *44*(2), 118-134.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. London, England: Temple Smith.

Smith, M. K. (2001). Chris Arygris: Theories of action, double-loop learning and organizational learning. *The Encyclopaedia of Informal Education*. Retrieved April 25, 2011, from www.infed.org/thinkers/argyris.htm.

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 433-449.

Tee, M. Y., Samuel, M., Norjoharudden, M. N., & Nadarajan, S. (2016). A methodological approach for researching national classroom practices. *Journal of International and Comparative Education*, *5*(1), 1-17.

Thanh-Pham, T. H., & Renshaw, P. (2014). Formative assessment in Confucian heritage culture classrooms: Activity theory analysis of tensions, contradictions and hybrid practices. *Assessment & Evaluation in Higher Education*, *40*(1), 45-59.

Thomas, L., Deaudelin, C., Desjardins, J., & Dezutter, O. (2011). Elementary teachers' formative evaluation practices in an era of curricular reform in Quebec, Canada. *Assessment in Education*: *Principles, Policy & Practice*, *18*(4), 381-398.

Willis, D. (1993). Learning and assessment: Exposing the inconsistencies of theory and practice. *Oxford Review of Education*, *19*(3), 38-402.

Willis, J. (2011). Affiliation, autonomy and Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 399-415.

Wiliam, D. (2006). Assessment: Learning communities can use it to engineer a bridge connecting teaching and learning. *Journal of Staff Development*, *27*(1), 16-20.

Wiliam, D. (2010). *Embedded formative assessment*. Bloomington, USA: Solution Tree.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Mahwah, USA: Lawrence Erlbaum Associate.