

3L: *The Southeast Asian Journal of English Language Studies* – Vol 25(4): 209 – 226
<http://doi.org/10.17576/3L-2019-2504-13>

The Validation of *Assessment for Learning* Audit Instrument: A Mixed Methods Approach

MAZIDAH MOHAMED

Universiti Kebangsaan Malaysia
mazidah78@gmail.com

MOHD SALLEHHUDIN ABD AZIZ

Universiti Kebangsaan Malaysia

KEMBOJA ISMAIL

Universiti Kebangsaan Malaysia

ABSTRACT

The CEFR-aligned curriculum promotes integration in the implementation of Assessment for Learning (AfL) by English language teachers in primary schools. This paper depicts the mixed methods pilot and validation of the Assessment for Learning Audit Instrument (AfLAI) for the purpose of examining the use and understanding of AfL among the teachers in primary schools. In Phase 1, three subject matter experts were consulted, and nine English language teachers provided consensual validation. In Phase 2, 53 English language teachers responded to the adapted instrument. AfLAI was further discussed with the validators in Phase 3. This procedure was adapted from Phase 8 of the Instrument Development and Construct Validation mixed research techniques. Among the emerging themes were on language correction, item/ instrument difficulty, and useful items. The quantitative data was compared with the scales from the original authors regarding the four key strategies of AfL in AfLAI: Sharing Learning Intentions and Success Criteria, Questions and Classroom Discussions, Feedback and Peer- and Self-Assessment. Based on the mixed data arranged in a joint display, some parts in the items were adapted to accommodate the local context. Among the findings were (i) the key strategy known as Learning Intentions in AfLAI was not applicable in the target context, and (ii) some examples in the items needed to be changed to fit the Malaysian syllabus. The implications of this pilot study and validation process included 16 changes in the instrument and the translation of the instrument into Bahasa Melayu.

Keywords: Assessment for Learning; English language teachers; primary school; audit instrument; mixed methods

INTRODUCTION

The English Language Education Reform in Malaysia: 2015-2025 (Zuraidah et al. 2015), based on the Malaysian Education Blueprint 2013-2025 by the Ministry of Education, Malaysia (MoE) (2012), had included the discussion on the importance of achieving marketable capitals to meet the demands of the global economy (Hazita 2016). Based on the education system planning, the English Language Standard and Quality Council (ELSQC) under MoE Malaysia signed a 5-year contract with the Common European Framework of Reference (CEFR), beginning from 31 August 2016, to improvise the syllabus and enable the target language competency achievement according to the acknowledged framework. A national baseline assessment was administered online to a sample of Standard 6 students in 2012, and had provided some information for the determination of the minimum English language competency for primary school leavers, which is mid A2 or Basic Users (Nurul Farehah & Mohd Sallehhudin 2018). The improvisation of the target language among the pupils has to be made gradually based on the minimum requirement, and at the same time, allow those who excel to attain higher levels. Some issues in the English

language pedagogy, such as literacy and proficiency (Normazidah et al. 2012) could be tackled. Hence, the teachers were trained to implement the CEFR curriculum in scheduled cascades (Abdul Hakim Ali et al. 2018).

The CEFR-aligned English language curriculum was cascaded to all the English language teachers in Malaysian public schools and also some local-based private schools since 2017. The implementation of the CEFR-aligned curriculum began in 2018 (Abdul Hakim Ali et al. 2018). The teachers were trained to evaluate, adapt, differentiate and design suitable lessons for the learners based on the CEFR Curriculum Induction and Learning Materials Adaptation courses. The suggested strategies on asking questions and differentiation were also included in the cascading of the curriculum, and these are parts of Formative Assessment (FA) and Assessment *for Learning* (AfL).

In 2018, the cascade on Standard 3 syllabus had reinforced the implementation of FA. FA has been emphasized since the inauguration of School-Based Assessment (SBA) in 2011, marking the beginning of the increased workload issue (Azlin et al. 2013). The revision of FA in 2018 had further connected the teachers to the implementation of AfL as a part of classroom assessment. The cascade included 9 building blocks for FA: Sharing learning objectives and success criteria, exemplars, starters and plenaries, deliberate practice, questioning, discussions, quick scans, self-assessment and peer-assessment, and feedback, among others, by quoting Wiliam (2018). In this study, four key strategies of AfL are piloted and validated based on the Assessment *for Learning* Audit instrument (AfLAI) by Lysaght & O’Leary (2013): Sharing Learning Intentions and Success Criteria, Questions and Classroom Discussions, Feedback and Peer- and Self-assessment.

ISSUES IN THE IMPLEMENTATION OF *ASSESSMENT FOR LEARNING*

FA has been implemented in public Malaysian classrooms since 2011, in line with the SBA policy (Alla Baksh et al. 2016), along with Summative Assessment (SA). Alla Baksh et al. (2016) focused on the washback effects of SBA on the students, and later on the pre-service teachers in secondary schools (Alla Baksh et al. 2019). Within FA, Assessment *for Learning* (AfL), Assessment as Learning and Assessment of Learning are also implemented.

Previously, Rohaya, Mohd Zaki, Hamimah and Adibah (2014) surveyed on the Malaysian teachers’ competency in AfL and the result showed low understanding among the respondents. Perhaps the low understanding on AfL is caused by limiting the definition of assessment as pencil and paper tests and high stakes examinations per se. When an embedded assessment such as FA or AfL is introduced, some educators were less willing to pause and delve deeper into the suggested content, even though they implemented it. In addition to that, the stigma on added workload due to FA and the assessment evidence continues (Nurfaradilla et al. 2010).

A more recent study by Abdul Hakim Ali et al. (2018) pinpointed on the cascade of the CEFR-aligned curriculum, but they did not focus on the implementation of AfL. Moreover, the article also voiced concern that the new CEFR-aligned curriculum is far from being a perfect framework in the education system. There might not be a simple and fast solution to this issue, but being in the education system, educators as the implementers have the right to really study the situation under the guidance of inspectors of schools as mentors. Instead of finding faults, discussing for solutions could make the burden lighter.

There is a saying, ‘if you do what you always did, you will get what you’ve always gotten’. As the new policy and the revised curriculum are introduced, there is a responsibility to learn and unlearn new practices among all the levels in the education system. In a webinar, the Director of Education advised all the officers in the ministry to embrace change, and not just implement the changes in the curriculum (Amin 2019). Only then the changing of a perspective towards a better practice can be appreciated by all the officers, regardless whether they are the policy makers, the quality management officers or the implementers. The new curriculum is not meant to be a surface change per se. The top-down curriculum may not satisfy all the requests from the current stakeholders, but it has been written based on a standard framework – CEFR.

Unfortunately, since the high stakes examination is still functioning as the most sought-after test and the most important result, the change in balancing the implementation of FA and SA becomes very difficult to practice. The washback effect: ‘teaching to the test’ may not be easily be reviewed and reapplied as ‘using summative tests formatively’. Research on FA sometimes discourage the implementers from participating due to the added workload that has little to do with higher achievement in high stakes examination. This study offers a slight difference whereby the implementation of FA and SA are integrated in the education system provided that the 4 key strategies in AfL are used parallel to the other innovations such as 21st Century skills and Higher Order Thinking Skills. Hence, the gap of studies in the implementation of AfL in the Malaysian context was critically assessed as one of the issues that need further investigation.

On the other hand, numerous studies have reported on the implementation of FA in other countries. Despite being good resources, Black (2015), criticized the articles on FA whereby the data collected were mostly from self-reports, and not rich data from real interactions in the classroom. A whole volume of the *Assessment in Education: Principles, Policy and Practice* journal was dedicated on the reports on the implementation of FA back in the year 2015. As Lysaght & O’Leary (2013) and Black (2009, 2015) pinpointed, the data from self-reports must be judged carefully as they may not be perfectly accurate (Lysaght 2009; Ryan et al. 2012) because as humans, we tend to be positive and choose to show our best foot forward.

The generic issue in FA and AfL is the skyrocketing workload among the teachers. One of the articles that support this notion was from Singapore, where Ratnam-Lim & Tan (2015) reported on the tediousness of the implementation of FA through detailed written feedbacks, et cetera. In the beginning of the PERI-HA (Primary Education Review and Implementation for Holistic Assessment) in Singapore, summative assessment was excluded from Primary 1 and Primary 2 classes. It was found that lengthy written feedback was not easily comprehended by the young learners.

Similar in Malaysia, the issue of overwhelming workload occurred in the beginning of the implementation of SBA (Azlin et al. 2013). When the MoE Malaysia abolished standardized SA for Standard 1, Standard 2 and perhaps Standard 3 students in 2019, the stigma of increased workload persists. Despite the similarity, there is also a difference whereby Malaysian teachers are given more empowerment in choosing the more suitable AfL strategies based on their diverse students (Mazidah & Mohd Sallehhudin 2018). However, the Curriculum Development Centre and the MoE Malaysia still provided some templates and forms for the teachers to report on Classroom Assessment (Pentaksiran Bilik Darjah). None of the forms mentioned FA or AfL. Based on AfLAI, teachers are not required to fill student assessment forms all the time, instead, the focus is on the implementation of the AfL key strategies in the classrooms. Previous implementations of FA that required too many forms and checklists resulted in the “tick box

syndrome” (Marshall & Drummond 2006) and it is hoped that the case would not repeat in Malaysia.

In order to carry out a survey on the teachers’ use and understanding of AfL, the AfLAI by Lysaght & O’Leary (2013) was adapted via interviews, pilot study and discussions on the local version of implementation of AfL. The purpose of this validation process is to prepare an instrument that can be related to the situations faced by the English language teachers in primary schools, related to the implementation of AfL in an on-going study.

ASSESSMENT FOR LEARNING AUDIT INSTRUMENT

AfL is not a process, nor a limited set of procedures. There may not be a “one-theory-fits-all” or a special model in AfL yet. One of the conclusions by Black & Wiliam (1998) on AfL is it is an ongoing process which should include all the pupils in the classroom, and each individual teacher has their own way to practice the strategies in their own unique style.

In other words, the implementation of AfL depends on the individual styles by the teachers, plus their interactions with the pupils, and does not demand numerous forms as evidence. Becoming too dependent on forms and reports on paper may someday affect the original plan in teaching and instructions which should emphasize more on interactions with the learners. The target of including AfL in the comprehensive model of pedagogy is to prove how a child could become the evidence of his/her own progress, and not the evidence from added paper work per se, despite the fact that a human child can be unpredictable most of the time. The key word is interaction. Strategies in interaction during pedagogy could lead to self-regulated learning strategies among the learners (Philip & Tan 2006).

Leahy et al. (2005) defined AfL as being aware of “where are we now, where are we going, and how to get there”. In a more recent discussion, Wiliam (2018) described FA as having evidence to decide on the next instructions in pedagogy.

This discussion brings to the notion that the best evidence may be in the unpredictable child itself, not in the forms. There is also the emphasis on nurturing the passion for learning among young learners, and moving towards self-regulated learning as the aim of the implementation of AfL.

Next, to audit the interaction as the evidence, AfLAI has to be adapted and validated. This validation process highlights the phases in capturing the implementation of AfL that requires various considerations with regards to the local culture and the real situations in the classrooms. Lysaght & O’Leary (2013, 2017) actually used the AfLAI in a comprehensive Continuous Professional Development (CPD) and Teacher Learning Community (TLC) projects for the participants. Lysaght (2009) began her study in a disadvantaged school and focused on the CPD to relate the implementation of AfL among the teachers and the young learners’ reading achievement. Even though there was no significant difference between the reading achievement between the control and experimental groups of young learners in her study, it was found that the teachers’ attitude and classroom practices did change (Lysaght 2009).

MoE has also introduced Professional Learning Community (PLC) in Malaysia. Unlike in Malaysia, the Irish education system had introduced bottom-up curriculum and syllabus in classrooms 10 years before Lysaght & O’Leary’s (2017) study (Department of Education Ireland 2007). Therefore, there is a need to adapt the instrument to match with the current classroom situation and culture in Hulu Langat. Perhaps the classroom culture is influenced by instructional

leadership as in the assessment regarding LINUS 2.0 English language screening in Sibulahi, whereby efficient headmasters contribute to the teachers and learners achieving 100% literacy rate in their respective schools (Sio & Ramlee 2019). Or perhaps, by using AfLAI, the good pedagogical practices among the teachers could be documented.

The 58 items in AfLAI are considered as a more holistic description on the teachers' understanding of AfL. There was also another instrument known as Assessment *for* Learning Measurement instrument (AfLMi) (Lysaght et al. 2013) which had “reduced” the 58 items in AfLAI into 20 items only. But the authors of AfLAI did not recommend using AfLMi in order to examine the teachers' understanding of the key strategies of AfL.

The original instrument listed Learning Intentions and Success Criteria, Questioning and Classroom Discussions, Feedback and Self- and Peer-Assessment as the key strategies of Assessment *for* Learning (Lysaght & O'Leary 2013, 2017). There are two versions of AfLAI, one is the 2013 version, the other is the 2017 version. There are only 5 rating scales in the latest version of AfLAI. Previously, (Lysaght & O'Leary 2013) distributed the instrument with 6 rating scales – “embedded”, “established”, “emerging”, “sporadic”, “never” and “do not understand”. One was omitted in the 2017 version. The 6th rating scale/response: “Do not understand” was taken out from the latest AfLAI in the year 2017 based on the Irish teachers' beliefs (Lysaght & O'Leary 2017). TABLE 1 shows the list of the key strategies of AfL and the rating scales in AfLAI.

TABLE 1. Key Strategies and Rating Scales in AfLAI

4 Key strategies of AfL (58 items)	6 Rating scales	Explanation
Learning intentions and success criteria (16 items)	Embedded	Happens 90% of the time
Questioning and classroom discussion (16 items)	Established	Happens 70% of the time
Feedback (12 items)	Emerging	Happens 50% of the time
Peer- and self-assessment (14 items)	Sporadic	Happens 25% of the time
	Never	Never happens
	Do not understand	I do not understand what the statement means

Upon granting permission to use and adapt the instrument, the authors suggested using the 2017 version, but due to some feedback from the validators, the 2013 version was used in this pilot study and validation process. The following subtopic explains more on the methods applied in this validation process.

METHODOLOGY

PARTICIPANTS

The target population of this on-going study is 772 English language teachers in 89 public primary schools within Hulu Langat district. To adapt the original instrument, AfLAI went through qualitative phases of validation by consulting 3 subject matter experts (SMEs) and 9 purposively sampled teachers. The English language teachers as validators were purposively sampled from the target population, and it was also concept sampling to represent the implementers of AfL (Merriam 2002, 2009; Patton 1990, Creswell 2012:208) The qualitative findings in the first phase and the third phase in this study were collected from the same validators. The rationale for using the teachers as validators is because they can provide consensual validation into the audit instrument. Eisner (2017:112) defined consensual validation

as “...agreement among competent others that that the description, interpretation, evaluation and thematic of an educational situation are right.” The pilot instrument was also distributed to 70 teachers by hand via convenient sampling (Creswell 2012), and 53 responded.

TABLE 2 shows the background of the validators. Some of these validators also gave follow-up responses in the third phase which is also qualitative. From the 9 validators, 1 was a male, 2 were Masters’ degree holders, 2 had more than 25 years of teaching experience, and 4 had less than 10 years of teaching experience. Only two of them were teaching Level 1 pupils. All of them had been invited to attend the CEFR-aligned English language curriculum cascade at school level. In this cascade, the teachers were required to study the notes on the implementation of the CEFR-aligned curriculum. They were also given softcopies of a handbook with the 9 building blocks of FA. In the cascade trainer notes, the teachers were instructed to do micro- and macro-teaching sessions using the CEFR-aligned curriculum Content Standards and Learning Standards. Along with the documents, the textbooks in the CEFR-aligned curriculum were also very comprehensive. Provided that the teachers follow, adopt and adapt the teacher guidebooks, almost all the activities in the CEFR-aligned textbooks contain suggestions on FA and A/L, along with other 21st Century Skills prescribed by the MoE Malaysia. In a nutshell, the validators and all the English language teachers in Malaysia should have a certain degree of exposure and experience in the implementation of FA and A/L, especially level 1 teachers within the new CEFR-aligned curriculum.

TABLE 2. Background of validators

Validator ID	Gender	Age	Teaching experience (years)	Academic background
V1	F	42	10	Bachelor
V2	F	39	10	Bachelor
V3	F	55	31	Bachelor
V4	F	33	8	Master’s
V5	F	35	8	Bachelor
V6	F	33	8	Master’s
V7	F	40	11	Bachelor
V8	F	34	8	Bachelor
V9	M	48	26	Diploma

Meanwhile, the respondents in the quantitative phase are shown in TABLE 3. There were 53 respondents, with four males. Most of the respondents were between 30 to 39 years old, with less than 10 years of teaching experience. At least 9 respondents were Master’s degree holders. Most of them were Bachelor degree holders. Four of the respondents were more than 50 years old and indicated that their highest academic achievement as SPM holders, but they have more than 30 years of teaching experience. These teachers actually graduated from Teachers’ College and received a teaching certificate. Moreover, these senior teachers have experienced more than 3 changes in the curriculum – *Kurikulum Baru Sekolah Rendah*, *Kurikulum Bersepadu Sekolah Rendah*, *Kurikulum Standard Sekolah Rendah* and *Kurikulum Standard Sekolah Rendah 2017* review.

Five of the 53 teachers were teaching special needs classes. The rest of them were teaching the mainstream classes. Overall, the validators consisted of 92% female, 50% were between 30-39 years old, 43% were Level 1 teachers, and 75% were degree holders. All the public primary schools have mainstream classes, but only selected school open the special needs classes. In the original study, Lysaght and O’Leary stated that 30% of the respondents were

special educators, and there were no significant differences between the special educators and mainstream educators. TABLE 3 shows the background of the respondents in Phase 2.

TABLE 3. Background of respondents in Phase 2

Total Respondents	Gender	Age	Academic background	Teaching experience (years)
53	Male (4)	20 – 29 (9),	SPM (4)	0 – 9 (23)
	Female (49)	30 – 39 (27),	Bachelor (40)	10 – 19 (22)
		40 – 49 (11)	Master's (9)	20 – 29 (5)
		50 and above (6)		30 years and above (3)

PROCEDURE

This mixed methods process adapted the *Instrument Development and Construct Validation* (IDCV) by Onwuegbuzie, Bustamante, & Nelson (2010) based on Phase 8 out of the suggested 10 phases. IDCV is used for the purpose of optimizing the development of quantitative instruments and it comprises of multiple research frameworks, models and approaches in the form of a meta-framework (Onwuegbuzie, Bustamante & Nelson 2010: 56). Since this is a mixed-method study, the data is presented in a joint display (Creswell 2014: 84) in the discussion session. The importance of adapting the IDCV method is to prepare a valid modified instrument for a larger sample in the population of the English language teachers in primary schools.

In the first phase, the AfLAI instrument was adapted based on the awareness of the local culture in the implementation of AfL, by considering cross-cultural adaptation of research instruments (Gjersing et al. 2010). Originating from Ireland, the instrument may or may not be totally suitable in the Malaysian context. The idea of AfL and FA may be similar across countries, but the interpretation and the implementation surely needed some revisions and adaptations.

The IDCV method in instrument development was adapted to revise AfLAI and to make it relatable to the implementation of AfL in Hulu Langat. There are 10 phases in IDCV, however, only 1 phase from IDCV was included in the validation of the adapted AfLAI, which was Phase 8: Validate Revised Instrument: Mixed Analysis Phase: Qualitative-Dominant Crossover Analyses (Onwuegbuzie et al. 2010). The reason for adapting only Phase 8: the qualitative-dominant crossover analyses was this study used an instrument that has been developed and validated by the original authors.

This study applied qualitative-dominant crossover analyses (Onwuegbuzie, Bustamante & Nelson 2010, p. 61) based on the use of qualitative methods in the 1st and 3rd phase of validating AfLAI in the Hulu Langat context. The first phase was audiotaped and transcribed interview with 9 participants, the second phase was the distribution of AfLAI to 70 respondents, and the 3rd phase was written interview, whereby the same participants in the first phase wrote their comments on the twice adapted AfLAI.

The first phase used qualitative interviews. Nine validators discussed about the instrument and gave their opinions in verbal and in writing. The interviews were transcribed broadly, analysed and some themes were identified and tabulated.

The second phase was quantitative survey whereby 53 out of 70 English language teachers responded to the instrument via convenience sampling (Creswell 2012). Data from the second phase was analysed using SPSS using factor analysis and Cronbach's alpha reliabilities.

In the third phase, the data was collected via another qualitative phase that was written interviews with the purposively selected respondents in the first phase (Merriam 2009; Patton 2014).

Both the qualitative and quantitative data were analysed, discussed, then integrated and displayed in the discussions section. The following section describes the findings and the results.

FINDINGS AND RESULTS

Phase 1. During in the first process of qualitative validation, the 2013 version of AfLAI was used. The instrument was lengthy with 58 items. The sociodemographic part in the instrument was adapted from an INTO (Irish National Teachers Union) survey that had 14 semi-structured item focusing on school assessment. Apparently, some of the validators were daunted by the 58 items in AfLAI due to the SBA related questions in the sociodemographic part. Eventually, the INTO questions in the sociodemographic part were eliminated and replaced with 14 structured questions.

The subject-matter experts also suggested a review of top-down documents to find the instructions from MoE Malaysia regarding the implementation of AfL in Formative Assessment. The suggested strategies to implement AfL which were incorporated in AfLAI (Self- and Peer-Assessment, Feedback, Classroom Questionings and Discussions, Learning Objectives and Success Criteria) was finally documented in the 2017 KSSR review in the English Language *Standards-Based Curriculum and Assessment Document DSKP*, which was implemented in 2019 as a part of Formative Assessment (MoE Malaysia 2018). TABLE 4, shows the qualitative findings.

TABLE 4. Some themes from Phase 1

Themes	Written feedback	Verbal feedback
Language correction	(Corrected the grammar in item 53) (V1)	V1 commented on a number of grammar corrections needed. The term Assessment <i>for</i> Learning was deemed ungrammatical.
Item difficulty	‘Some items may be hard to understand’ (V2)	(Maybe the key strategies of AfL are) “...not applicable because of the level of the students...” (V4)
Term difficulty	‘Components – to simplify’ (V4) ‘Avoid examples from other countries’ (V4) ‘Terms – should make the participants become aware of the meanings’ (V4) (Asked for examples on item 14) (V5)	Asked, “what is the difference between assessment <i>for</i> learning and of learning?” (V9)
Useful for peer coaching (in PLC)	‘Can be used for peer-coaching’ (V8) ‘Good for teachers to understand’ (V7) ‘Examples are helpful’ (V2) (These items) ‘cover the normal practice in classrooms’ (V2)	None

During Phase 1, 6 validators gave written feedback and 3 validators commented on the instrument. The whole discussion took two hours and the validators were given flexible time when they were free to attend the session at their convenience. Among the identified themes in Phase 1 were on: Language correction; Item difficulty; Term difficulty and Useful for peer coaching (in PLC).

Phase 2. After the qualitative phase, the INTO survey in the socio-demographic section was omitted. One of the key strategies of AfL was changed into Learning Objectives and Success Criteria (LOSC) from Learning Intentions and Success Criteria (Lysaght & O’Leary 2013). The other key strategies: Questioning and Classroom Discussion (QCD), Feedback (FB) and Peer- and Self-Assessment (PSA) were retained. The themes found in Phase 1 were also reviewed and some examples were improvised.

TABLE 5 shows the reliability and factor analyses data from the second phase of the instrument validation. The data from the pilot was compared with the original AfLAI by Lysaght & O’Leary (2013). The reason for comparing this result with the 2013 version was because this validation study used 6 rating scales. There was no comparison with the 2017 version of the instrument because the more recent version by Lysaght & O’Leary incorporated only 5 rating scales after omitting the: ‘I do not know’, which is the 6th scale. However, based on this result, the 2017 version of AfLAI would be used in this on-going study because there was almost no response indicating: ‘I don’t know’.

TABLE 5 Comparison of results in Phase 2

	Lisc	Losc	Qcd ¹	Qcd ²	Fb ¹	Fb ²	Psa ¹	Psa ²
Number of items	16	16	16	16	12	12	14	14
Cronbach’s Alpha Reliability	0.92	0.85	0.89	0.90	0.83	0.85	0.88	0.94
Factor 1 Eigenvalue	7.2	5.5	6.1	6.8	4.4	4.6	5.5	8.6
Per cent of variance explained	45.1	34.5	38.6	42.7	36.6	38.6	39.5	61.5

*1 = from (Lysaght & O’Leary 2013), *2 = from pilot study

From TABLE 5, the Cronbach’s Alpha Reliability shows that the validated instrument is acceptable and can be considered satisfactory. No item is removed from the adapted instrument.

Phase 3. Earlier in Phase 1, the validators had suggested that some phrases in the original AfLAI were not understood and not relatable to the Hulu Langat classroom culture, such as the term: *Learning Intentions*. This term was changed into Learning Objectives. Moreover, there were many other questions from the validators regarding the terms in the instrument. This issue was unavoidable as the terms sometimes had overlapping meaning or referred to the activities implemented by the teachers, but they were not familiar with the names of the activities. Responding to these questions, the SMEs suggested that the instrument needed to be translated into Bahasa Melayu. Sometimes, the use of mother language could provide support because learning and using a second language is not an easy task (Noor Hayati & Mohd Sallehudin 2015).

DISCUSSIONS

Overall, adopting mixed methods approaches to validate AfLAI was an invaluable experience in order to collect and portray the data from different perspectives. The issues in the Mixed Methods Research (MMR) as discussed by Timans, Wouters & Heilbron (2019:213) called for a reflection on the efforts of the early MMR researchers who may have been able to ‘enhance a specific (contextual) form of knowledge’, in this case, the validation of the implementation of

AfL strategies by English language teachers in the classrooms. Details on the validation of the AfL/Ai in mixed methods data are as the following subtopics.

SHARING LEARNING OBJECTIVES AND SUCCESS CRITERIA

The results for the LOSC strategy. The following lists the AfL practices reported by teachers as being the most embedded/established in their own classrooms. In the original study, the most embedded practice was item 5 which stated the sharing of the learning intentions using child-friendly language (mean=5.26) (Lysaght & O’Leary 2013). But in this pilot study, item 11 was the most practiced strategy whereby assessment techniques are used to assess pupils’ prior learning via concept mapping, and iThink map (mean=4.90). The least practiced strategy in both original study and this pilot study was item 8: “Prompts are used to signal LOSC with pupils (e.g. using WALTs (We Are Learning Today...) and WILFs (What I’m Looking For...) in lower primary classes” (means=3.29 (Lysaght & O’Leary 2013), 3.61). In this study, the respondents were not familiar yet with the WALTs and the WILFs.

QUESTIONS AND CLASSROOM DISCUSSIONS

Responses regarding QCD showed that the teachers usually use hinge questions more during lessons compared to asking pupils about the purpose of a lesson. In the original study, the most embedded practice was item 3: “Questions are used to elicit pupils’ prior knowledge on a topic” (mean=5.44) (Lysaght & O’Leary 2013). However, item 3 was ranked as the second most embedded (mean=5.16). The most embedded practice in QCD was item 4, whereby the teachers claimed they used hinge questions to determine pupils’ progress in lessons (mean=5.30). The least embedded item was item 8: “Assessment techniques are used to encourage questioning of the teacher by pupils (e.g. using parking lot, Q & A session, hot-seating or a Post-its challenge)” (mean=3.37) (Lysaght & O’Leary 2013). On the contrary, the least embedded practice in this pilot study was item 16: “Pupils are asked to explain why they are undertaking particular tasks (e.g. the teacher might ask, ‘why are we completing this worksheet/what are we learning by doing it?’)” (mean=4.14).

FEEDBACK

Next, the teachers’ self-reports on Feedback strategy. The most embedded practice in the original study was item 1 which states “Feedback is given according to the original Learning Intentions and Success Criteria” (mean=4.82) (Lysaght & O’Leary 2013). In this study, item 1 was ranked at the fourth out of the 12 statements. The most embedded practice was item 4: “Teachers’ praise of pupils’ work (e.g. ‘that’s excellent; well done’) is deliberately and consistently supplemented with Feedback that specifies the nature of the progress made” (mean=5.19). The least embedded practice was item 7: “Pupils are involved formally in providing information about their learning to their parents/guardians” (mean=4.15).

PEER- AND SELF-ASSESSMENT

Last but not least, the ranking of PSA in the local context showed item 1 as the most embedded practice whereby: “Pupils are given an opportunity to inform what they feel about the topic, activity or lesson, in the beginning of the session (e.g. using traffic lights)” (mean=4.68). In the

original study, it was item 3: “Lessons on new topics begin with pupils being invited to reflect on their prior learning” (mean=4.42) (Lysaght & O’Leary 2013). The least embedded practice was item 2: “Pupils are encouraged to record their progress using, for example, learning logs” (mean=3.56). Apparently item 12 was more embedded in this study (mean=4.02) compared to the original study whereby “Time is set aside during parent/guardian-teacher meetings for pupils to be involved in reporting on some aspects of their learning” (mean=2.48) (Lysaght & O’Leary 2013).

COMPARISON OF AfLAI SCALES

TABLE 6 shows the overall average rating for each of the four scales in rank order in this study, labelled as hypertext 2. The data is also compared with the original study, labelled as hypertext 1. The highest rating was on the practice of Questions and Classroom Discussions, which is different from the original study which stated Sharing Learning Intentions and Success Criteria as the most practiced strategy (Lysaght & O’Leary 2013). A one-way repeated measured analysis of variance (ANOVA) was conducted to evaluate and confirm that the mean ratings for the AfL strategies were statistically significantly different ($F(3, 156) = 173.021, P < 0.0005$) and the effect size difference was large ($\eta^2 = .96$). Post-hoc tests, using the Bonferroni correction, revealed that the mean for all the scales were not statistically significantly different ($P < 0.0005$ in all cases).

All 58 items from the four scales had mean teacher-ratings of 3.5 or more, suggesting that all the key strategies of AfL were fairly understood and used in their respective classrooms. This is different from the original study whereby their practice of PSA was sporadic with the rating of 3.5 or lower in most items (Lysaght & O’Leary 2013). TABLE 6 shows the differences between the original study and this pilot study.

TABLE 6. Comparison of the AfLAI scales

	Mean	SD	
QCD ¹ /QCD ²	4.2 ¹ /4.6 ²	0.80 ¹ /0.83 ²	Emerging
FB ¹ /FB ²	4.2 ¹ /4.6 ²	0.73 ¹ /0.84 ²	Emerging
Sharing LISC ¹ /LOSC ²	4.4 ¹ /4.4 ²	0.68 ¹ /0.93 ²	Emerging
PSA ¹ /PSA ²	3.3 ¹ /4.1 ²	0.74 ¹ /0.95 ²	Sporadic ¹ /Emerging ²

*1 = from (Lysaght & O’Leary 2013), *2 = from pilot study

VALIDATION OF AfLAI IN A JOINT DISPLAY

Based on the overall AfLAI scales and the feedback from the participants, the following table shows the data integration. This data integration display is arranged according to a *building into a qualitative instrument display* (Creswell 2014) and *data consolidation* (Onwuegbuzie et al. 2010).

AfLAI could either make the respondents discouraged from the language, the items and the terms in the instrument due to the difficulty (V1, V2, V4 & V5); or spark improvised ideas (V8, V7 and V2). For example, it was questioned: why would teachers implement AfL vigorously when what matters most is high stakes results in the Primary School Achievement Test (UPSR) by the end of primary schooling. This issue is identified as the washback effect of high stakes examinations (Alla Baksh et al. 2016; Yahya et al. 2014). Only one finding in Phase 1 could be regarded as a ‘spark’, which was the theme: AfL is useful. According to V8, the key strategies of AfL could be used in peer coaching in PLC.

In TABLE 7, the first column portrays the parts of the instrument. The second column shows the adapted pilot instrument parts in the AfLAI. The third column shows the feedback frequencies regarding the First Phase of piloting the instrument. The changes made by the researcher is labelled as ‘Etic’, and the changes suggested by the original authors and the SMEs as ‘Emic’. In the third column, the means of the scales are shown.

Based on the qualitative and quantitative feedback, 16 changes were made to the AfLAI. The following table shows the data integration in a joint display:

TABLE 7. Validation of AfLAI instrument display

Instrument parts	Adapted pilot instrument	Feedback frequency	Mean	Changes made
Demographic section	INTO Survey	1	-	14 questions only
Rating scale	Embedded	Emic	-	I practise this all the time
	Established			I practise this often
	Emerging			I seldom practise this
	Sporadic			I rarely practise this
	Never			I never practise this
	Do not understand			Omitted (Lysaght & O’Leary 2017)
Sharing LOSC Item L1, L2, L4, L6, L12, L13, L14, L15, L16, F43	Learning Intentions...	6	4.88, 4.76,4.49, 4.16, 4.53, 4.55, 4.71, 3.88, 4.04, 4.49	Learning Objectives...
Item 8	... junior classes	6	3.61	... primary classes
Item 11	E.g. concept mapping...	Emic	4.90	E.g. concept mapping, iThink mind map, formative assessment, summative assessment
QCD Item 20	E.g. we have been learning to sort 3D shapes that stack and roll. Now, if you were given a choice, would you build a tower with spheres or cubes?	Emic	5.16	E.g. we have been reading about Mowgli in The Jungle Book. Which adjective that describe Mowgli do you like most? Why?
Item 24	E.g. using hot seating or a Post-it challenge	6, Emic	4.50	E.g. using parking lots or Q & A sessions
Item 27	... using think-pair-share, for example	6, Emic	4.70	... using think-pair-share, group discussions, presentations, for example
FB Item 34	E.g. thumbs-up-thumbs-down and/or two stars and a wish	6, Emic	4.85	e.g. thumbs-up-thumbs-down, two stars and a wish and/or traffic light
Item 35	Written feedback	1, Emic	5.02	Written and/or spoken feedback
Item 37	E.g. identifying common mistakes in the addition of fractions	Emic	4.74	E.g. identifying common mistakes in spelling
Item 38	E.g. common errors in the comprehension section of the MICRA-T... (etic)	Emic	4.62	E.g. common errors in the comprehension section
Item 39	E.g. portfolios or learning logs are taken home	6, Emic	4.15	E.g. portfolios or learning logs are taken home, corrections are written
Item 41	Closing-the-gap-feedback...	6, Emic	4.17	- formative feedback...
Item 42	... uses a variety of prompts...	6, Emic	4.49	... uses a variety of strategies...
PSA Item 45	... to indicate how challenging, they anticipate the learning will	6, Emic	4.68	To inform what they feel about the topic, activity or lesson...

These were some of the mixed methods findings arranged in a joint display (Creswell 2014, p. 84). The means of the items are compiled in the appendix. The feedback in the qualitative phases are integrated in column three. Meanwhile the fourth column is based on the responses from 53 participants in the quantitative phase. The numbers of the items were also changed for the purpose of labelling the data as it was difficult to differentiate the items from the different key strategies. The changes were made to the AfLAI instrument for the purpose of distribution in the on-going field study. The following section is a summary of the validation of the adapted AfLAI instrument process.

CONCLUSION

This study intended to validate and pilot an available instrument in order to capture the implementation of AfL by the English language teachers in primary schools within the local context and classroom culture. Some of the adaptations that have been decided were (i) the omission of the INTO survey in the socio-demographic section, (ii) the changes in 36 terms, and (iii) the instrument is translated into Malay language. Despite a recent argument by Brown (2019) about AfL strategies not being a valid assessment method, he acknowledged the strategies as suggestable for pedagogical interaction in a classroom full of humans. Interactions with inputs and outputs have been found to increase the vocabulary knowledge of young learners (Chin et al. 2019), thus promoting learning. The mixed methods approach adapted in this validation process supports the rationale of ‘conducting preliminary exploration with individuals to make sure that instruments actually fit the participants and site being studied’ (Creswell 2014:15). It is hoped that the adapted instrument in the on-going study could connect the teachers to the key strategies of AfL and contribute by bringing out the best practices that can be developed into a practical toolkit on the implementation of AfL in primary schools.

REFERENCES

- Abdul Hakim Ali, A. A., Radzuwan, A. R. & Wan Zhafirah, W. Z. (2018). The enactment of the Malaysian Common European Framework of Reference (CEFR): National master trainer’s reflection. *Indonesian Journal of Applied Linguistics*. 8(2), 409–417. doi:10.17509/ijal.v8i2.13307
- Alla Baksh, M. A. K., Mohd Sallehudin, A. A. & Siti Hamin, S. (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), 51–68.
- Alla Baksh, M. A. K., Mohd Sallehudin, A. A., Yahya, A. T. & 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.
- Amin bin Senin (Ketua Pengarah Pelajaran Kementerian Pendidikan Malaysia). (2019). Naratif baharu amalan pendidikan. Fokus & Gerak Kerja Profesional KPPM 2019. *BICARA PROFESIONAL KPM*. EduwebTv. <https://www.youtube.com/watch?v=vTLk1F7A9uA>
- Azlin, N. M., Ong, H. L., Mohamad Sattar, R., Rose Amnah, R. & Nurhayati, Y. (2013). The benefits of school-based assessment. *Asian Social Science*. 9(8), 101–106. doi:10.5539/ass.v9n8p101
- Black, P. (2009). Formative Assessment issues across the curriculum: the theory and the practice. *TESOL Quarterly*. 43(3), 519–524. doi:10.1002/j.1545-7249.2009.tb00248.x
- Black, P. (2015). Formative Assessment - an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*. 22(1), 161–177. doi:10.1080/0969594X.2014.999643
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy &*

- Practice*. 5(1), 7–74. doi:10.1080/0969595980050102
- Brown, G. T. L. (2019). Is Assessment for Learning really assessment? *Front. Educ.* 4(64): 1–7. doi:10.3389/educ.2019.00064
- Chin, M. L. L., Krishnamoorthy, K. & Yap, J. R. (2019). The role of negotiated interaction in L2 vocabulary acquisition among primary ESL learners. *3L: The Southeast Asian Journal of English Language Studies*. 25(2), 1–21. doi:http://doi.org/10.17576/3L-2019-2502-01
- Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 4th Edition. Pearson.
- Creswell, J. W. (2014). *A Concise Introduction to Mixed Methods Research*. SAGE Publications.
- Department of Education Ireland. (2007). The Northern Ireland Curriculum: Primary. *CCEA*.
- Eisner, E. W. & Noddings, N. (2017). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. Early childhood education series. Teachers College Press. Retrieved from <https://books.google.com.my/books?id=09UcDgAAQBAJ>
- Gjersing, L., Capplehorn, J. R. M. & Clausen, T. (2010). Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Medical Research Methodology*. 10(13), 1–22. doi:10.1186/1471-2288-10-13
- Hazita, A. (2016). Implementation and challenges of English language education reform in Malaysian primary schools. *3L: The Southeast Asian Journal of English Language Studies*. 22(3), 65–78.
- Leahy, S., Lyon, C., Thompson, M. & Wiliam, D. (2005). Classroom Assessment: minute by minute, day by day. *Educational Leadership*. 63(3), 19–24.
- Lysaght, Z. (2009). *From Balkanisation to boundary crossing: using a Teacher Learning Community to explore the impact of assessment on teaching and learning in a disadvantaged school*. Unpublished PhD thesis, Dublin City University, Ireland.
- Lysaght, Z. & O’Leary, M. (2013). An instrument to audit teachers’ use of assessment for learning. *Irish Educational Studies*. 32(2), 217–232. doi:10.1080/03323315.2013.784636
- Lysaght, Z. & O’Leary, M. (2017). Scaling up, writ small: using an assessment for learning audit instrument to stimulate site-based professional development, one school at a time. *Assessment in Education: Principles, Policy & Practice*. 24(2), 271–289. doi:10.1080/0969594X.2017.1296813
- Lysaght, Z., O’Leary, M. & Ludlow, L. (2013). Measuring teachers’ Assessment for Learning (AfL) classroom practices in elementary schools. *International Journal of Educational Methodology*. 3(2), 103–115. doi:10.12973/ijem.3.2.103
- 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. doi:10.1080/02671520600615638
- Mazidah, M. & Mohd Sallehuddin, A. A. (2018). Juxtaposing the primary school assessment concepts and practices in Singapore and Malaysia. *International Journal of Engineering & Technology*. 7(3.21), 552–556.
- Merriam, S. B. (2002). *Qualitative Research in Practice: Examples for Discussion and Analysis*. The Jossey-Bass higher and adult education series. Jossey-Bass. Retrieved from <https://books.google.com.my/books?id=exe2AAAAIAAJ>
- Merriam, S. B. (2009). *Qualitative Research: A Guide to Design and Implementation*. Higher and adult education series. John Wiley & Sons. Retrieved from <https://books.google.com.my/books?id=tvFICrgcuSIC>
- Ministry of Education Malaysia Malaysia Education Blueprint 2013-2025 (Preschool to Post-Secondary Education) (2012). Retrieved from http://www.moe.gov.my/cms/upload_files/articlefile/2013/articlefile_file_003108.pdf
- Ministry of Education Malaysia. (2018). Dokumen Standard Kurikulum dan Pentaksiran Bahasa Inggeris Tahun 3 Sekolah Kebangsaan. Putrajaya. Retrieved from <http://bpk.moe.gov.my/index.php/terbitan-bpk/kurikulum-sekolah-rendah>
- Noor Hayati, R. & Mohd Sallehuddin, A. A. (2015). The use of Bahasa Melayu in the English language classroom by ‘non-optionist’ English teachers. *Procedia - Social and Behavioral Sciences* 172(2015): 770–777. doi:10.1016/j.sbspro.2015.01.431
- Normazidah, C. M., Koo, Y. L. & Hazita, A. (2012). Exploring English language learning and teaching in Malaysia. *GEMA Online Journal of Language Studies*. 12(1), 35–51.
- Nurfaradilla, N., Siti Norhidayah, R., Mohammad Iskandar, S., Kasmah, A. B. & Sharifah Nor, P. (2010). Teachers’ perception on alternative assessment. *Procedia - Social and Behavioral Sciences*. 7(C), 37–42. doi:10.1016/j.sbspro.2010.10.006
- Nurul Farehah, M. U. & Mohd Sallehuddin, A. A. (2018). Implementation of CEFR in Malaysia: Teachers’ awareness and the Challenges. *3L: The Southeast Asian Journal of English Language Studies*. 24(3), 168–

183.

- Onwuegbuzie, A. J., Bustamante, R. M. & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*. 4(1), 56–78. doi:10.1177/1558689809355805
- Patton, M. Q. (2014). *Qualitative Evaluation & Research Methods*, 4th Edition. Thousand Oaks, CA, US: Sage Publications, Inc.
- Philip, B. & Tan, K. H. (2006). Metacognitive Strategy Instruction (MSI) for reading: co-regulation of cognition. *Jurnal e-Bangi*. 1(1), 1–27.
- Ratnam-Lim, C. T. L. & Tan, K. H. K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment in Education: Principles, Policy & Practice*. 22(1), 61–78. doi:10.1080/0969594X.2014.1001319
- Rohaya, T., Mohd Zaki, K., Hamimah, A. N. & Adibah, A. L. (2014). From principle to practice: Assessment for Learning in Malaysian school-based assessment classroom. *International Journal of Social Sciences and Education*. 4(4), 8. Retrieved from <http://ijsse.com/sites/default/files/issues/2014/v4-i4-2014-1/Paper-11.pdf>
- Ryan, K., Gannon-Slater, N. & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*. 33(3), 414–430. doi:10.1177/1098214012441499
- Sio, J. J. L. & Ramlee, I. (2019). Binary logistic regression analysis of instructional leadership factors affecting English language literacy in primary schools. *3L: The Southeast Asian Journal of English Language Studies*. 25(2), 22–37. doi:<http://doi.org/10.17576/3L-2019-2502-02>
- Timans, R., Wouters, P. & Heilbron, J. (2019). Mixed methods research: what it is and what it could be. *Theory and Society*. 48(2), 193–216. Retrieved from https://link.springer.com/article/10.1007/s11186-019-09345-5#Fn7_source
- Wiliam, D. (2018). *Embedded Formative Assessment*, 2nd Edition. Solution Tree Press.
- Yahya, A. T., Mohd Sallehuddin, A. A., Kemboja, I. & Alla Baksh, M. A. K. (2014). The washback effect of the General Secondary English Examination (GSEE) on teaching and learning. *GEMA Online Journal of Language Studies*. 14(3), 83–103. doi:10.17576/GEMA-2014-1403-06
- Zuraidah, M. D., Anna Christina, A., Arshad, A. S., Mardziah Hayati, A., Singh, K. K. K., Lee, B. H., Pillai, J. P. @ L., et al. (2015). *English Language Education Reform in Malaysia: The Roadmap 2015-2025*. Putrajaya: Ministry of Education Malaysia.

APPENDIX

TABLE 8. Average ratings for the Sharing LOSC scale: rank ordered by practices that are most and least embedded

Sharing LOSC	N	Mean	SD
11. Assessment techniques are used to assess pupils' prior learning (e.g. concept mapping, iThink mind map, formative assessment, summative assessment...).	52	4.90	0.84
1. Learning Objectives are shared with pupils at appropriate times during lessons (e.g.: Halfway through the lesson, the teacher might say: 'remember, we are learning to describe a friend').	53	4.88	0.78
5. Child-friendly language is used to share Learning Objectives with pupils (e.g. 'We are learning to make a good guess (prediction) about what is likely to happen next in the story').	53	4.88	0.88
3. Pupils are reminded about the links between what they are learning and the big learning picture (e.g. 'We are learning to count money so that when we go shopping, we can check our change').	53	4.78	1.00
2. Learning Objectives are stated using words that emphasise knowledge, skills, concepts and/or attitudes i.e. what the pupils are learning NOT what they are doing.	53	4.76	0.96
14. Pupils' progress against the key objectives of the lessons is noted and/or recorded as part of lessons (e.g. in the reflections in the lesson plan).	53	4.71	0.91
10. Samples of work are used to help pupils develop a nose for quality.	53	4.65	0.69
13. Learning Objectives are available throughout lessons in a manner that is accessible and meaningful for all pupils (e.g. written on the black/whiteboard and/or in pictorial form for Lower Primary classes).	53	4.55	1.04
12. Pupils are reminded of the Learning Objectives during lessons.	53	4.53	0.76
4. Pupils are provided with opportunities to internalise and express Learning Objectives by, for example, being invited to read them aloud and/or restate them in their own words.	53	4.49	1.21
9. Success Criteria are differentiated for according to pupils' needs (e.g. the teacher might say, 'Everyone must complete parts 1 and 2...; some pupils may complete part 3').	53	4.35	0.77
6. Success Criteria related to Learning Objectives are differentiated and shared with pupils.	52	4.16	0.98
16. Pupils are given responsibility for checking their own Learning Objectives and Success Criteria of lessons.	53	4.04	0.95
7. Pupils are involved in identifying Success Criteria.	52	3.92	0.88
15. Pupils demonstrate that they are using Learning Objectives and/or Success Criteria while they are working (e.g. checking their progress against the Learning Objectives and Success Criteria for the lesson displayed on the blackboard or flipchart, for example).	53	3.88	1.09
8. Prompts are used to signal Learning Objectives and Success Criteria with pupils (e.g. using WALTs (We Are Learning Today...) and WILFS (What I'm Looking For...) in Lower Primary classes).	52	3.61	1.07

TABLE 8. Average ratings for the QCD scale: rank ordered by practices that are most and least embedded

QCD	N	Mean	SD
4. During lessons, hinge questions are used to determine pupils' progress in lessons (e.g. 'We have been learning to describe our friends using the words 'long hair', 'tall' and 'black hair'. What are other words that we can use to describe them?')	52	5.30	0.70
3. Questions are used to elicit pupils' prior knowledge on a topic.	52	5.16	0.79
2. Assessment techniques are used to facilitate class discussion (e.g. brainstorming).	52	4.94	0.81
6. Assessment techniques are used to encourage all pupils to engage with questions (e.g. no hands up, names out a hat, etc.).	51	4.82	0.77
5. Assessment techniques are used to activate pupils/get them thinking during discussions and/or questioning (e.g. using think-pair-share or talk partners).	52	4.76	0.74
1. When planning lessons, key, open-ended questions are identified to ensure that pupils engage actively in lessons (e.g. 'If we close all the windows in the classroom, do you think the wind and the sunlight can come inside?').	52	4.72	0.90
11. Pupils are asked to explore their own ideas with others, using group discussions, presentations and think-pair-share, for example.	53	4.70	0.90

13. Pupils' incorrect responses are used to guide teaching and learning (e.g. a pupil is asked to explain why he/she gave a particular answer).	52	4.68	0.86
8. Assessment techniques are used to encourage questioning of the teacher by pupils (e.g. using parking lot, Q & A session, hot-seating or a Post-its challenge).	53	4.50	0.73
10. The pace of class discussions is deliberately slowed down to encourage all pupils to think before responding (e.g. using wait time).	53	4.50	0.99
9. Questioning goes beyond the one right answer style (where the focus is often on trying to guess the answer in the teacher's mind) to the use of more open-ended questions that encourage critical thinking.	53	4.48	0.86
7. Pupils are encouraged to share the questioning role with the teacher during lessons (e.g. the teacher routinely invites pupils to question their peers' contributions to discussions).	52	4.42	0.88
14. Pupils are asked to evaluate their peers' responses to questions (e.g. 'Aina, do you agree with what Melinda has said and why?').	53	4.34	0.89
12. Individual answers to questions are supplemented by pupils taking an answer 'round the class so that a selection of responses from the pupils is used to build a better answer'.	53	4.32	0.81
15. Pupils can explain to others what they are learning (e.g. if a visitor came to the classroom, pupils could articulate what they are learning in terms that identify the knowledge, skills and/or attitudes being developed).	53	4.16	0.76
16. Pupils are asked to explain why they are undertaking particular tasks (e.g. the teacher might ask, 'why are we completing this worksheet/what are we learning by doing it?').	53	4.14	0.80

TABLE 9. Average ratings for the FB scale: rank ordered by practices that are most and least embedded

Feedback	N	Mean	SD
4. Teachers' praise of pupils' work (e.g. 'that's excellent; well done') is deliberately and consistently supplemented with Feedback that specifies the nature of the progress made (e.g. 'Well done Zetty, this paragraph helps me to visualise the characters in the story because of the adjectives you use').	53	5.19	0.73
3. Written feedback on pupils' work goes beyond the use of grades and comments such as 'well done' to specify what pupils have achieved and what they need to do next.	53	5.02	0.72
2. Assessment techniques are used during lessons to help the teacher determine how well pupils understand what is being taught (e.g. traffic lights, thumbs-up-thumbs-down and/or two stars and a wish).	53	4.85	0.81
1. Feedback to pupils is focused on the original Learning Objective(s) and Success Criteria (e.g. 'Today we are learning to use punctuation correctly in our writing and you used capital letters and full stop correctly in your story, well done Amir').	53	4.75	0.78
5. Teacher-made tests are used diagnostically to identify strengths and needs in teaching and learning (e.g. identifying common mistakes in writing email closures).	53	4.74	0.96
6. Diagnostic information from standardised tests is used to identify strengths and needs in teaching and learning (e.g. common errors in the comprehension section Paper 1 are identified and used in teaching, or, Paper 2, the Writing examination is discussed).	53	4.62	0.92
10. When providing feedback, the teacher goes beyond giving pupils the correct answer and uses a variety of prompts/strategies to help them progress (e.g. scaffolding the pupils by saying: 'You might need to use some of the new adjectives we learned last week to describe the characters in your story').	53	4.49	0.82
8. Feedback focuses on one or two specified areas for improvement at any one time (e.g. in correcting written work, punctuation errors may not be marked if the primary focus of the writing is on the use of adjectives).	53	4.43	0.72
12. Pupils are provided with information on their learning on a minute-by-minute, day-by-day basis rather than end of week/month/term.	53	4.36	0.83
9. Closing-the-gap feedback is used to focus pupils' attention on the next step in their learning.	53	4.17	1.03
11. In preparing to provide pupils with Feedback on their learning, the teacher consults their records of achievement against key Learning Objectives from previous lessons (e.g. the teacher reviews a checklist, rating scale or anecdotal record that s/he has compiled).	53	4.17	0.80
7. Pupils are involved formally in providing information about their learning to their parents/guardians (e.g. corrections, portfolios or learning logs are taken home or discussed during academic open day).	53	4.15	0.92

TABLE 10. Average ratings for the PSA scale: rank ordered by practices that are most and least embedded

PSA	N	Mean	SD
1. Pupils are given an opportunity to inform what they feel about the topic, activity or lesson, in the beginning of the session (e.g. using traffic lights).	52	4.68	0.91
3. Lessons on new topics begin with pupils being invited to reflect on their prior learning (e.g. pupils complete a mind map or concept map or brainstorm a topic).	51	4.46	0.88
6. Pupils are encouraged to use a range of assessment techniques to review their own work (e.g. a rubric, traffic lights, thumbs up/down, two stars and a wish).	51	4.30	0.93
4. Pupils are provided with opportunities to reflect on, and talk about, their learning, progress and goals.	51	4.14	0.92
11. Pupils use each other as resources for learning (e.g. response/talk partners who comment on each other's work and discuss how it can be improved).	51	4.12	0.82
9. Assessment techniques are used to create an environment in which pupils can be honest about areas where they are experiencing difficulty (e.g. talk partners are used to facilitate conversations between pupils about the challenges they face in their learning).	51	4.10	0.97
5. Pupils assess and comment on each other's work (e.g. they are taught how to use the success criteria of a lesson to judge another pupil's piece of work).	51	4.06	0.86
7. A visual record of pupils' progress is maintained to celebrate pupils' learning and show areas of/for development (e.g. a bulletin board displaying progression in story writing over a term).	51	4.02	0.95
12. Time is set aside during parent/guardian-teacher meetings for pupils to be involved in reporting on some aspects of their learning (e.g. pupils select an example of their best work for discussion at the meeting).	51	4.02	1.07
8. Time is set aside during lessons to allow for Self- and Peer-Assessment.	51	3.94	0.89
13. Pupils use differentiated success criteria to Self- and/or Peer-Assess (e.g. pupils can distinguish between what must be achieved to be successful on a task and what might be done to gain extra credit).	51	3.90	0.99
14. Pupils have ready access to exemplar materials showing work at different levels of achievement across a range of subject areas (e.g. pupils use examples of collage and the Art display board when advising peers on how to improve their work).	51	3.84	0.95
10. When pupils have difficulty in their learning, they are encouraged to draw on a range of self-assessment strategies and techniques to help them overcome the problem (e.g. they consult with an exemplar on the bulletin board).	51	3.76	0.77
2. Pupils are encouraged to record their progress using, for example, learning logs.	53	3.56	1.31