Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989 – 9572

DOI: 10.47750/jett.2022.13.02.006

Construct Validity Analysis: Assessment for Learning Primary Mathematics Questionnaire

Haiza Atmareni Harmeni[1]

Rohaya Talib[2]

Journal for Educators, Teachers and Trainers, Vol. 13 (2)

https://jett.labosfor.com/

Date of reception: 10 Dec 2021

Date of revision: 24 Feb 2022

Date of acceptance: 28 Feb 2022

Haiza Atmareni Harmeni, Rohaya Talib (2022). Construct Validity Analysis: Assessment for Learning Primary Mathematics Questionnaire *Journal for Educators, Teachers and Trainers,* Vol. 13(2). 57 – 75.

[1]PhD Student(Measurement and Evaluation in Education)School of Education,Faculty of Social Science and Humanities,Universiti Teknologi Malaysia
[2]Associate Professor,School of Education,Faculty of Social Science and Humanities,Universiti Teknologi Malaysia

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 13 (2)
ISSN 1989 – 9572

https://jett.labosfor.com/

Construct Validity Analysis: Assessment for Learning Primary Mathematics Questionnaire

Haiza Atmareni Harmeni[1],Rohaya Talib[2]

[1]PhD Student(Measurement and Evaluation in Education)School of Education,Faculty of Social Science and Humanities,Universiti Teknologi Malaysia

[2]Associate Professor,School of Education,Faculty of Social Science and Humanities,Universiti Teknologi Malaysia

Email: atmareni@gmail.com,rohayatalib@utm.my

ABSTRACT

Construct validity is an essential aspect of developing an instrument that involves the collection of evidence to justify decisions. A pilot study was conducted in Hulu Langat, Selangor, to identify the construct validity of a questionnaire developed to measure the practice of Assessment for Learning among primary school Mathematics teachers in the Central Zone, Malaysia. This study applied a simple random sampling technique involving 406 teachers who agreed to perform self-assessment using an online questionnaire. Data analysis was performed using AMOS 23 software. The findings showed only 55 out of 72 items achieved the fit for Confirmatory Factor Analysis. Next, the 12 constructs studied met the statistical requirements for Discriminant Validity and Composite Reliability. The findings of this study are discussed, and suggestions for improvement are given. The construction of a valid and reliable instrument contributes to the study of measurement and evaluation. This contribution meets the needs and fills the research gap in psychometrics.

Keywords: Assessment for Learning, Primary School Mathematics, Construct Validity, Convergent Validity, Discriminant Validity, Composite Reliability

INTRODUCTION

The National Education Policy (NEP) 2017 was refined to ensure that everyone involved can achieve their educational aspirations. The National Philosophy of Education (NPE) 1988 was the foundation of this national policy. A guide for all educators relating to the strategies implemented to achieve this goal was the Malaysian Education Blue Print (MEBP) 2013-2025. School quality and student accomplishment are two essential indicators of the Malaysian educational system's success. Students would get meaningful learning when developing knowledge, skills, and values. So, Malaysian education has to focus on development-oriented learning according to the needs of 21st Century Learning. Teachers need to assess students continuously during teaching and learning (TnL) in the classroom.

The Education Ministry (MoE) extended the Classroom Assessment (CA) to level one students in primary schools, including years one, two, and three in 2019. CA, formerly known as School-Based Assessment (SBA), has been implemented in Malaysia since 2011. In 2016, SBA was known as CA due to its implementation procedures continuously in TnL to obtain information on development, progress, abilities and student achievement. There are two approaches used in CA such as formative and summative. Assessment for Learning (AfL) and Assessment as Learning (AaL) are formative, while Assessment of Learning (AoL) is summative. The application of AfL on the constructivist approach in TnL Mathematics has positively impacted students' mastery of problem-solving skills (Chemeli, 2019).

Mathematics is one of the best mediums to develop intellectual potential and professionalism. Due to its innate nature, human development promotes logical and systematic thinking that can contribute to forming students who have a Mathematical and holistic mindset as intended in NPE. This mathematical-minded student is a student who can do mathematics and understand mathematical ideas as well as responsibly apply mathematical knowledge and skills in daily life based on mathematical attitudes and values. Mathematical thinking also aims to produce creative and innovative individuals and meet the needs of the 21st century because the country's ability is highly dependent on human capital that can generate creative and new ideas. Emphasis on the developmental aspects of students' thinking mathematically is built and developed through the TnL process in the classroom based on problem-solving, communication, reasoning, relevance, representation making, and the use of technology in mathematics (MoE, 2017).

TnL is beyond teaching students basic literacy, but teachers should guide how students learn (Teo, 2019). Instructional design and approach are critical factors in supporting students to achieve standards and provide opportunities to demonstrate the effectiveness of AfL (Heritage & Wylie, 2018). Teachers could apply 4C skills (communication, creative, critical, collaborative) and relate to real-world life in line with the rapid development of technology (Anagün, 2018). A conducive learning environment promotes competition among students. A good learning environment could foster competencies, motivate students to engage in various activities, and recognise the values and attitudes. Constructivist learning practitioners believe that students build knowledge and learn actively through interaction with the environment. At the same time, teachers play a role as facilitators to guide students to create meaningful understanding (Peter & Stout, 2006).

Nevertheless, teachers stated that AfL is a complex process because teachers need to have the ability to diagnose, observe and analyse the learning process (Akhmedina, 2017). In Malaysia, teachers face difficulties understanding the assessment guidance documents (Kadir, Nor & Hutagalung, 2019). It isn't apparent because the guidance given is too subjective and causes teachers to implement AfL based on their understanding and perception (Jamil & Said, 2019) in different ways (Arumugham, 2018). Teachers do not integrate AfL throughout the TnL process (Varatharaj, Abdullah & Ismail, 2015) but only assess the final stage of teaching (Arumugham, 2020). Teachers are more likely to conduct individual assessments by making the Level of Mastery (LM) the primary indicator for implementing AfL (Arumugham, 2020). Such practices are contrary to the main principles of AfL, which prioritise the developmental process of student learning over achievement (Danielson, 2017).

In conclusion, AfL is not implemented effectively in Malaysia. According to the theory, there is no detailed guidance causing teachers to be unsure of how the practices meet the prerequisites and face challenges in practice (Arumugham, 2018). Thus, a questionnaire named AfLPMQ was developed to assess AfL practices and determine the most influencing items contributing to the AfL practices among primary school Mathematics teachers in the Central Zone of Malaysia. The AfLPMQ needs to undergo a statistical analysis to ensure that the data obtained have good psychometric characteristics and high validity and reliability. Therefore, this study was conducted to identify the construct validity and reliability for the AfLPMQ.

ASSESSMENT FOR LEARNING

Assessment for learning, also known as formative assessment, is one form of assessment for the constructivist approach. Duckor & Holmberg (2019) stated that formative assessment is the process of teachers and students giving feedback to adapt movements and continuous learning tactics. Effectively implemented formative assessments could help students to achieve targeted learning outcomes. Therefore, teachers should assess their students during learning to obtain information on student progress and understanding. Teachers need to adapt accordingly based on the collected evidence (Asamoah, 2019).

According to Lyon, Nabors & Wylie (2019), teachers need to define materials and identify students' levels of knowledge, abilities, and experience to plan appropriate learning and facilitation activities. While Mohamed, Kamis, & Ali (2016) emphasise students' learning styles be adapted to the learning components of knowledge, skills, and values. Bezabih, Yigzaw & Ahmed (2019) stated that teachers should also refer to the syllabus and handbooks provided to identify the learning standards that need to be mastered by students. Then only, it is structured into clear and specific learning goals (Mohamed et al., 2016). Finally, teachers need to form assignments that meet learning requirements (Rahimi & Sahragard, 2019). The delivery of learning content should also be done with the help of appropriate tools and materials by integrating technology in TnL for sharing, material exploration, and assessment.

During the learning process, teachers are advised by Kenyon (2019) to ask questions related to previous learning and connect them to what will be learned (Ozan & Kincal, 2018). Teachers also need to explain what is expected and why students learn (Mohamed et al., 2016). Transmitting knowledge to learning content (Lyon et al., 2019), teachers have to use appropriate analogies or examples (Khan & Hussain, 2019) to encourage students to develop their comprehensiveness (Gan, He & Mu, 2019). In addition, Ozan & Kincal (2018) suggested that teachers use high-level questions to stimulate students' thinking and encourage students to discuss in groups and help other students (Rahimi & Sahragard, 2019). Ulfvarson, Oxelmark, & Jirwe (2018) also suggested that teachers should respond to students' conversations by making agreements such as "good," "excellent," or asking students to give applause.

Meanwhile, teachers also recommended evaluating the effectiveness of education based on evidence, such as hand gestures, observations, conversations, written exercises, and aloud reading techniques (Heredia, Furtak, Morrison & Renga, 2019; Mohamed et al., 2016). The observed student responses were analysed to identify students' strengths and weaknesses (Yidana & Anti Partey, 2018; Khan & Hussain, 2019). It further facilitates teachers to consider whether to change teaching methods (Yidana & Anti Partey, 2018), shift student group members (Ozan & Kincal, 2018), identify student needs (Bezabih et al., 2019), and provide learning mastery feedback to students as well as parents (DeLuca, Coombs & LaPointe-McEwan, 2019).

In summary, it can be concluded that AfL is a continuous assessment process in TnL that includes three phases: planning, implementation, and reflection. In the planning phase, teachers need to determine learning standards, provide learning resources and cultivate learning readiness among students. Teachers should share learning targets and success criteria during the implementation phase, apply questioning techniques, encourage active student involvement, and provide feedback and appreciation. Finally, for the reflection phase, teachers should use evidence of learning and interpret it for the quality of continuous learning. Thus, twelve dimensions of AfL were developed from this study to measure the assessment practices of primary school Mathematics teachers in the Central Zone of Malaysia, as shown in Table 1.

Table 1: Operational Definition

Dimension	Definition
Learning Standards (LS)	Teachers plan learning activities based on student's ability level by grading learning content according to difficulty level, using student-centred strategies and appropriate assessment methods in shaping Mathematical thinking by referring to curriculum standard documents.
Learning Resources (LR)	Teachers plan to share and integrate technology-assisted learning materials, the provision of interactive and entertaining training, the use of existing materials to explain the concepts of mathematics, and the use of appropriate mediums for early detection of student progress.
Learning Readiness (LRe)	Teachers identify students' existing knowledge related to the topic or skills learned to determine the appropriate learning style and relate the content to the real world to attract students to learn and utilise various levels of student achievement in applying leadership values.
Sharing Learning Targets and Success Criteria (SLT)	Teachers begin teaching by stimulating students' thoughts, sharing ideas, employing special formats to express ideas, and linking existing knowledge to learning skills by describing learning goals and success criteria.
Application of Questioning Techniques (AQT)	Equality of opportunity provided to students to engage in learning where the teacher stimulates high-level thinking with challenging questions includes space for students to think and ask questions and uses the information obtained to map students' thinking.
Active Involvement (AcI)	Teachers perform activities to encourage creativity among students, trigger critical thinking, and foster collaboration in completing tasks where communication in Mathematics process skills are emphasised to form character, personality, and national identity in students
Self and Peer Assessment (SPA)	Teachers check students' comprehension by using material cues while guiding students to assess their achievement by emphasising self-assessment of learning and other group work where justification of assessment should be given so that accuracy in Mathematical calculations can be fostered.
Feedback (Fb)	The teacher summarises the level of mastery, gives and obtains the student's consent, shows the correct answer in Mathematical calculations, and contrasts the student's responses to provide ideas for alternative solutions to improve the quality of the assignment.
Appreciation and Self-esteem (ASe)	Teachers' reinforcement skills aim to increase students' motivation to learn by rewarding based on achievement, encouraging healthy competition among students, expressing appreciation for students' efforts, providing encouragement, and recognising students' progress.
Learning Evidence (LEv)	Teachers present learning by examining students' suggestions and feedback, performing observations and evaluations of group assignments, and assessing students' level of mastery and active involvement in learning activities.
Analysing Information (AIn)	Analysis carried out on the achievement of learning objectives that help teachers identify students who have not mastered specific skills and errors of Mathematical concepts that are often made, the level of thinking and the way students think, and the affective needs of students.
Quality of Continuous Learning (QCL)	Teachers will implement continuous quality learning by re-teaching or using other methods to explain skills lacking, implement

	enrichment and remedial activities, involve parents in progress targets, and plan programs to enhance student excellence.

METHODOLOGY

This study focuses on developing a questionnaire instrument (AfLPMQ) to measure the practice of AfL among primary school Mathematics teachers in the Central Zone, Malaysia. The method of developing AfLPMQ refers to the four models proposed by Miller, Lovler, and McIntire (2015), Koy et al. (2017), Irwing and Hughes (2018), as well as Lele, Setiawan, and Sulhadi (2018). There are three phases involved in the AfLPMQ development process, namely conceptualisation, construction, and validation. The conceptualisation phase involved two steps: defining the concept and the construct. The construction phase also involved determining the instrument structure and constructing the items. Similarly, the validation phase also involved two steps; referring the experts and conducting a pilot study. The AfLPMQ development process had illustrated in Figure 1.

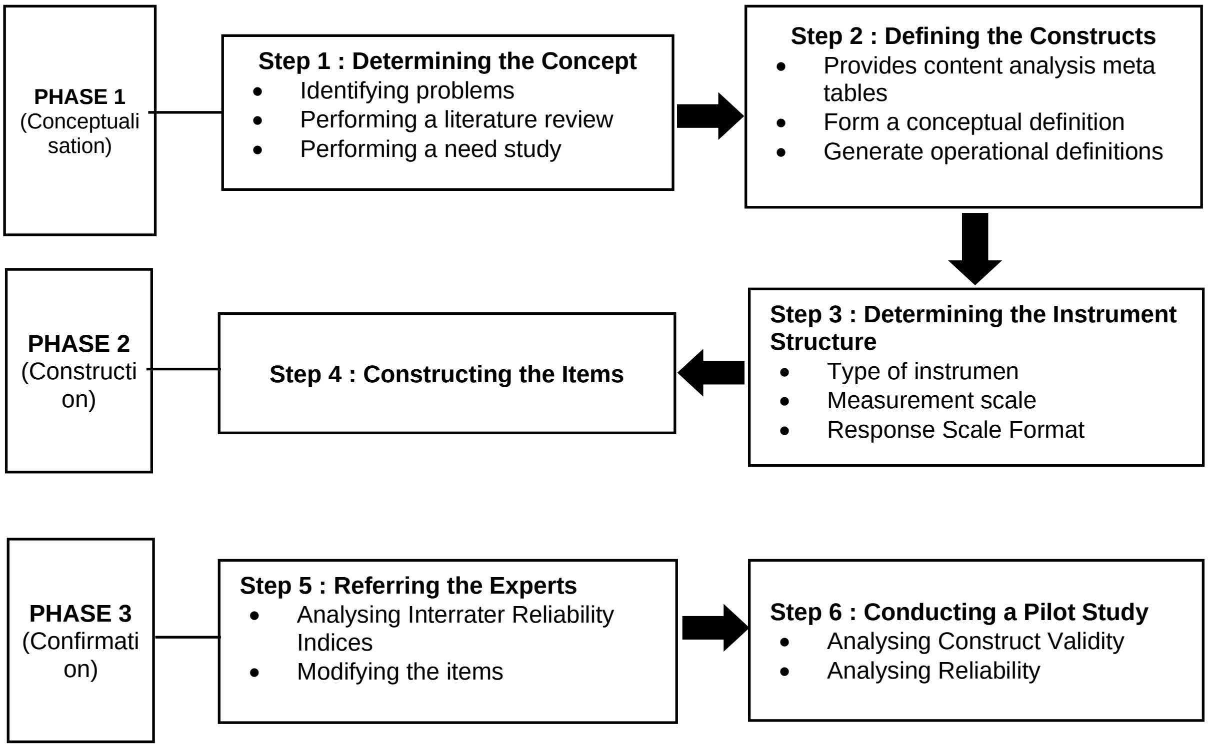

Fig.1:AfLPMQ development process

Phase 1: Conceptualisation

Step I, which determines the concept, referred to the 2019 Classroom Assessment Implementation Guidebook. Assessment for Learning (AfL) was selected as the idea studied after taking into account access to data, the benefits of the study on values, social and economic, and the contribution of knowledge in research as suggested by Leavy (2017). A literature review was also conducted by identifying the latest theories and studies related to AfL. The literature review also involves analysing reading material based on the definition of theory, the description of key concepts, methodology, findings, and relevance of sources with the study to be conducted.

The research gap has been identified based on the analysis of literature review conducted, and primary school Mathematics teachers in Malaysia's Central Zone (Federal Territory of Kuala Lumpur, Federal Territory of Putrajaya, and Selangor) were chosen as the population to be studied. Statistics obtained up to 30 June 2021 show that the number of Mathematics teachers is 4559 (Federal Territory of Kuala Lumpur = 895, Federal Territory of Putrajaya = 141, Selangor = 3523). So, a needs study was conducted through the interview method of three teachers in the Federal Territory of Kuala Lumpur, Putrajaya, and Selangor to find evidence of the phenomenon of research issues related to the AfL practices among Mathematics teachers in those states.

Step II is to define the construct. Guralnik (1976) represents a construct as an idea that results from synthesis. The approach proposed by Crocker and Algina (2008) is, Systematic Literature (SL) was used to define the construct of this study. This study applies the SL method of Literature Review, where the results of scholarly writing are obtained using the keywords Classroom Assessment, Formative Assessment, and Assessment for Learning. The reading materials received, such as articles, were screened based on the last four years, between 2016 to 2019. Only 23 papers were suitable to involve discussions on the concept and practice of AfL. Next, the idea is synthesised to formulate the construct and operational definitions. Detailed concept definitions and operational definitions help explain the constructs studied. Next, determining the dimensions and elements is done using the SL method of Mapping Study/Systematic Mapping, where the measurements and characteristics appropriate to the context of the study in Malaysia are selected.

Phase 2: Construction

Step III, which determines the instrument's structure, involves determining the type of instrument, measurement scale, format, and response scale. Miller and Lovler (2016) categorise psychological testing based on its purpose of either measuring cognitive, psychomotor or affective. Tests are more appropriate for measuring (cognitive) knowledge. Rubrics and observation checklists are commonly used to perform behavioural (psychomotor) observations. In contrast, the questionnaire measures self-assessment (affective). Thus, the questionnaire was chosen to be developed for this study and named AfLPMQ.

AfLPMQ contains two parts, namely part A and B. Part A is the respondents' demographic information. Three items are constructed in Part A, namely gender, academic qualifications, and teaching experience. Nominal scales were used for gender factors, while ordinal scales were used for academic qualification and teaching experience factors. Part B is information related to the practice of AfL among Mathematics teachers. The response format used is a five-point Agreement Type Likert Scale because it can be treated as an interval scale to calculate the mean, standard deviation, and other statistical tests (Sekaran and Bougie, 2016)

Step IV involves constructing items and providing instructions or instrument administration procedures to respondents. The first questionnaire constructed was named AfLPMQ Version 1. A total of 12 elements were developed. Each element contains six operational definitions. Thus, 72 items were constructed (six for each element) after considering the statistical requirement of at least three items to measure an element (Hair et al., 2019). The time taken to complete the assessment is also considered in determining the number of items to avoid fatigue among respondents to answer many items (Kline, 2005).

Phase 3: Validation

Step V, which refers to an expert, is implemented as soon as the AfLPMQ Version 1 items have been completed. A psychometric expert and two experts in the curriculum, pedagogy, and primary school mathematics assessment were appointed as evaluators of the AfLPMQ items. The expert panel was determined regarding the recommendations of Crocker and Algina (2008). Active involvement in conducting AfL and psychometric-related studies was the main criterion in selecting the expert panel. The three experts appointed are lecturers who have a Doctor of Philosophy degree and are currently serving at the Institute of Teacher Education, Malaysia. Consent from all three experts is obtained first before the letter of appointment is given.

The experts are responsible for evaluating each AfLPMQ item to ensure that it meets the content to be measured and meets the principles of item construction. They made an assessment for each item (2 = appropriate, 1 = needed to be modified, 0 = dropped). They also provided comments on less suitable items and suggestions for improvement for the modified items. The data obtained were analysed using Fleiss Kappa. Findings showed that out of 72 items, seven items needed to be modified (0.40 - 0.75), and two items needed to be dropped (<0.40). Nevertheless, AfLPMQ achieves content validity when the overall Fleiss Kappa Index of 0.93 is excellent (Fleiss, Levin & Paik, 1981). Thus, all nine items <0.75 were maintained with improvements implemented and named AfLPMQ Version 2.

Step VI is to perform a pilot study in which Instrument AfLPMQ Version 2 is tested to obtain the construct validity and reliability values. Pallant (2016) stated that the pilot study was aimed at; i) determining the reliability of the instrument, ii) estimating the time taken by the respondent to make the assessment, iii) ensuring the language used is easy to understand, iv) determining the validity of the construct, and v) determine the data collected can be analysed quantitatively.

This pilot study applies a quantitative research approach based on the problems and objectives of the study presented (Ali and Kerpčarová, 2019). A cross-sectional survey design is used because it involves many samples where data are collected at a given time to identify the practices being practised (Cresswell, 2012). This method is termed a list-based sampling method of high-coverage populations (Couper, 2000), which is applied following the Movement Control Order (MCO) currently in force to curb the spread of the Covid-19 pandemic. Data collection for this study was carried out online using a google form application sent to all schools via official email as soon as permission from the District Education Office was obtained.

Primary school Mathematics teachers in three states in the Central Zone of Malaysia, 4559 people are the target population for this study. Therefore, the simple random sampling technique was applied to determine the appropriate sample size. Thus, a total of two sources were consulted to determine the number of samples required. The sample size determination table (Krejcie and Morgan, 1970; The Research Advisor, 2006) suggests a total of 357 samples are needed for a population of 4559 with an error of 0.05%, while Hair et al. (2010) pointed out that one item requires a minimum of five respondents (1: 5). AfLPMQ contains 72 items, and the necessary respondents are 360 people. Cresswell (2012) suggested that the number of samples should be increased by 10% to accommodate the data dropped during data cleaning because it contains missing values and outliers. Pallant (2016) states that missing values are blank data, and outliers occur when entering data where numbers not within the scale are used. Collier (2020) suggested that data with a standard deviation value of less than 0.25 should be dropped because they are considered outliers that can cause mismatched data to produce a measurement model. So, a total of 400 samples was set as the minimum number after considering all the recommendations.

Construct validity is in the third phase, the final step (Step 6: Conducting a pilot study) before the instrument is used for data collection in the field (refer to Figure 1). Thus, a pilot study was conducted in Hulu Langat, Selangor, involving 519 Mathematics teachers from 91 primary schools. After data cleansing, only 406 teachers' data were analysed to test the validity of the construct statistically. The number of such samples is more than the prescribed minimum number and meets the requirements of factor analysis for construct validity requiring 300 samples (Tabachnick and Fidell, 2013).

Construct validity can be obtained through statistical analysis using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA analysis was used to determine the number of components underlying the elements studied (Hair et al., 2010). CFA analysis was intended to select the items constructed for each construct in line with the theory (Byrne, 2010). Nevertheless, EFA is not a prerequisite for CFA, as CFA can validate EFA findings (Phakiti *et.al*., 2018). Thus, this study only discusses the construct validity for AfLPMQ as evidenced through statistical analysis using CFA.

CFA is implemented one by one constructs through first-stage model analysis. At this stage, items that do not meet the Model Fit Index are dropped. The term "model fit" indicates that a specific model (estimated covariance matrix) represents data (observed covariance matrix). The AMOS software was used to analyse the CFA provided over 20 statistical tests for model fit. There are three types of Model Match Indices, namely Absolute Fit Indices (Chi-square probability level, GFI, AGFI), Incremental Fit Indices (TLI, NFI, CFI, RFI, ILI), and Parsimony Fit Indices (PNFI, RMSEA, CMIN/df). CFA model fit is determined by combining one of the indices from each type of Model Fit Index (Hu and Bentler, 1999; Hair *et al*., 2010).

In this study, the Maximum Likelihood Estimation method is used in the CFA exploration as proposed by Byrne (2010), where i) the internal adequacy of the model (t-value) is significant when the p-value <0.01, ii) the variance value does not exceed one or has a negative value, iii) the multiple correlation values for Standard Residual Covariances must be in the range of -2 to +2. P values (Absolute Fit Indices), NFI, RFI (Incremental Fit Indices), RMSEA, CMIN/df (Parsimony Fit Indices) were selected for reporting the Model Fit Index.

CFA is a good first step in determining construct validity, but CFA results alone have not proven Convergent Validity and Discrimination Validity (Collier, 2020). Convergent validity is a related analysis of a variable that measures only one construct (Widaman, 1985), while Discriminant Validity is an analysis to test whether a construct is related or different from another construct (Bagozzi *et al*., 1991). The framework outlined by Fornell and Larcker (1981) is the method used to determine Convergent Validity by calculating the value of Average Variance Extracted (AVE). The measured construct has Convergent Validity when the AVE value is higher than 0.50. Usually, AVE is also used as an indicator in determining Discriminatory Validity. Nevertheless, the decision to select the appropriate method is dependent on the results of the AVE calculations obtained. Therefore, Convergent Validity and Discriminant Validity are also discussed in this study.

RESULTS

Demography

A total of 406 teacher data were analysed, of which 120 (29.6%) were male, and 286 (70.4%) were female. The demographic analysis of the respondents is as in Table 2. As for the academic qualification factor, teachers with degrees are the most, which are 323 people (79.6%), followed by masters as many as 40 people (9.9%), STPM/Diploma as many as 28 people (6.9%), followed by SPM as many as 14 people (3.4) %) and only one (0.2%) teacher had a PhD. As for the teaching experience factor, the 11-20 years group is the most which are 202 people (49.8%), followed by the 1-10 years group which is 126 people (31.0%), 21-30 years group is 71 people (17.5%) and only seven teachers (1.7%) had more than 30 years of teaching experience.

Table 2:Demographics of respondents

Factors		Total	Percentage
Gender	Male	120	29.6
	Female	286	70.4
Academic Qualification	SPM	14	3.4
	STPM / Diploma	28	6.9
	Degree	323	79.6
	Masters	40	9.9
	PhD	1	0.2
Teaching Experience	1 – 10 years	126	31.0

	11 – 20 years	202	49.8
	21 – 30 years	71	17.5
	More than 30 years	7	1.7

Confirmatory Factor Analysis

The implemented first model analysis found that 17 of the 72 AfLPMQ were dropped based on recommendations obtained from the Modification Index to achieve model fit. The items dropped were SP37, SP61, SuP38, SuP50, KB15, KB51, BSP40, PTP29, PTP65, PA30, KRS19, KRS55, MB8, PHD33, EBP34, TM23, and TM47. The findings also show that 12 out of 55 items with a loading factor <0.50 are still acceptable when meeting the Measurement Index of the measurement model and are significant due to the large sample size (Hair *et al*., 2019). The result is shown in Table 3.

Table 3:Confirmatory Factor Analysis

Construct	CMIN	df	P	CMIN DF	NFI	RFI	PCFI	RMSEA	Item retained	
									Item No	Factor loading
LS	0.49	2	0.79	0.24	0.99	0.99	0.33	0.00	SP1	0.46
									SP13	0.46
									SP25	0.66
									SP49	0.49
LR	0.36	2	0.84	0.18	0.99	0.99	0.33	0.00	SuP2	0.61
									SuP14	0.63
									SuP26	0.56
									SuP62	0.37
LRe	1.22	2	0.54	0.61	0.99	0.99	0.33	0.00	KB3	0.41
									KB27	0.66
									KB39	0.61
									KB63	0.67
SLT	7.99	5	0.16	1.60	0.98	0.96	0.50	0.04	BSP4	0.57
									BSP16	0.60
									BSP28	0.67
									BSP52	0.50
									BSP64	0.59
AQT	2.01	2	0.37	1.01	0.99	0.98	0.33	0.00	PTP5	0.60
									PTP17	0.62
									PTP41	0.65
									PTP53	0.70
AcI	6.71	5	0.24	1.34	0.98	0.95	0.50	0.03	PA6	0.49
									PA18	0.34
									PA42	0.65
									PA54	0.70
									PA66	0.52
SPA	3.48	2	0.18	1.74	0.99	0.96	0.33	0.04	KRS7	0.68
									KRS31	0.18
									KRS43	0.87
									KRS67	0.52
Fb	6.03	5	0.30	1.21	0.98	0.96	0.50	0.02	MB20	0.45
									MB32	0.47
									MB44	0.50
									MB56	0.67
									MB68	0.70
ASe	9.70	5	0.08	1.94	0.98	0.95	0.49	0.05	PHD9	0.66
									PHD21	0.62
									PHD45	0.61
									PHD57	0.54
									PHD69	0.62
LEv	1.08	5	0.96	0.22	0.99	0.99	0.50	0.00	EBP10	0.62
									EBP22	0.65

									EBP46	0.73
									EBP58	0.67
									EBP70	0.55
AIn	3.87	2	0.14	1.94	0.98	0.95	0.33	0.05	TM11	0.31
									TM35	0.69
									TM59	0.67
									TM71	0.58
QCL	13.74	9	0.13	1.53	0.98	0.96	0.60	0.04	KPB12	0.72
									KPB24	0.61
									KPB36	0.64
									KPB48	0.67
									KPB60	0.50
									KPB72	0.49

Convergent Validity and Composite Reliability

All 55 items of AfLPMQ were retained after the first stage model analysis, then underwent the second-order model analysis where all elements were combined, as shown in Figure 2.

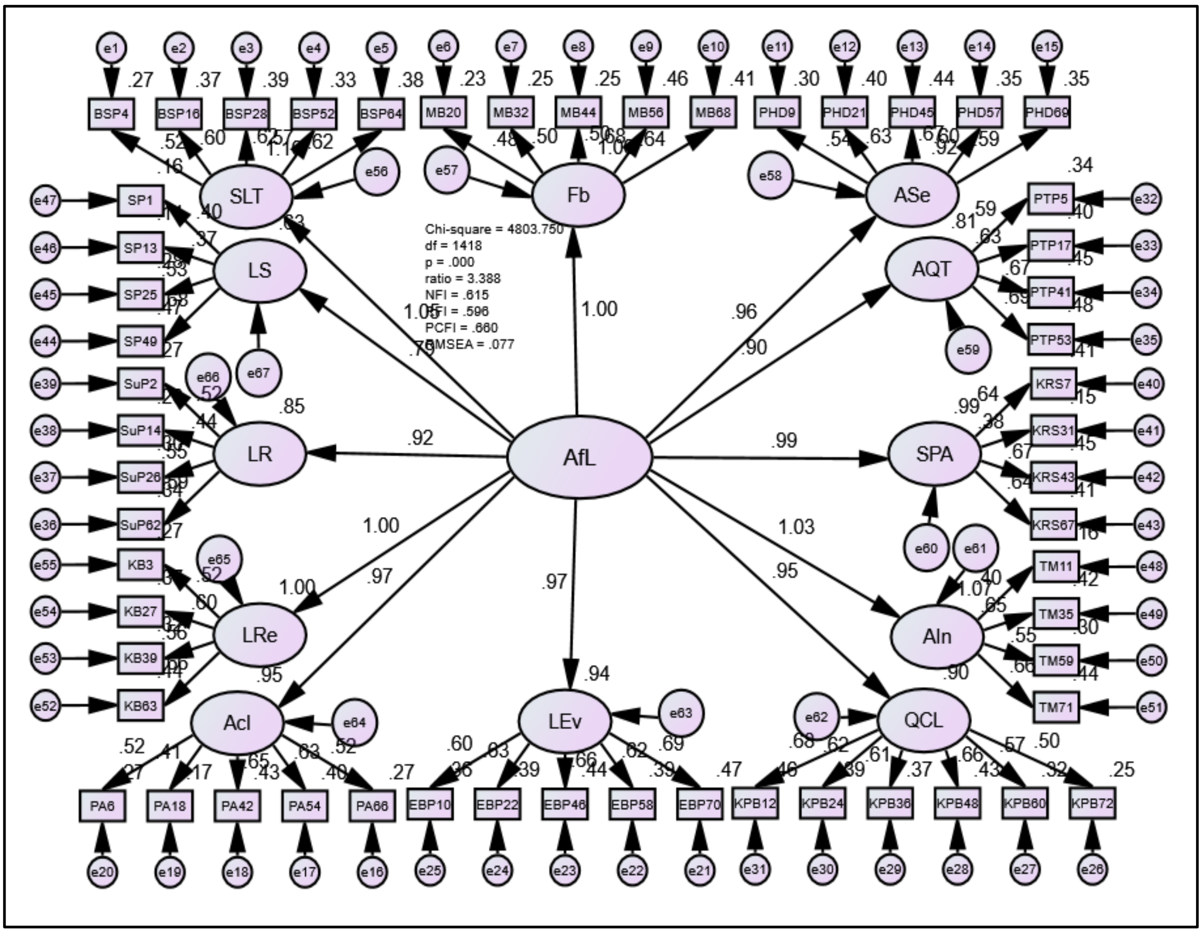

Fig.2:Second-order model analysis

Based on the analysis performed in Figure 2, the value of standardised factor loading (SFL) for each item is obtained. Subsequent analysis was performed using Microsoft Excel software. SFL values are squared (squared of standardised factor loading). The AVE value was obtained by calculating each construct's standardised factor loading squared average. The Composite Reliability (CR) value for each construct was also calculated based on the SFL value obtained. The Delta value for each item was first obtained by subtracting the SFL value from one (1 - SFL). Next, the delta values of all items for each construct are summed (sum of the delta).

Similarly, the value of SFL is summed (sum of SFL) and squared (squared of the sum of SFL). The sum of delta and square of the sum of the delta is added to form the denominator CR. The value of CR is obtained by dividing the squared value of the sum of SFL by the CR Denominator. The result found that all 12 elements studied had AVE values ranging from 0.26 to 0.42. The AVE value is less than 0.50 and less than meets Convergent Validity. However, the CR values obtained for all elements ranged from 0.66 to 0.85 and were at a reasonable and acceptable level because they exceeded 0.60 (Bond and Fox, 2015). The AVE and CR analyses are as in Table 4.

Table 4:Average Variance Extracted and Composite Reliability

Item	Standardised Factor Loading (SFL)	Squared of Standardised Factor Loading	AVE	Delta (1-SFL)	Sum of SFL	Squared of sum of SFL	Sum of Delta	CR Denominator	CR
SP1 <--- LS	0.40	0.16	0.26	0.60	1.98	3.93	2.02	5.95	0.66
SP13 <--- LS	0.36	0.13		0.65					
SP25 <--- LS	0.55	0.301		0.45					
SP49 <--- LS	0.67	0.45		0.33					
SuP2 <--- LR	0.53	0.28	0.28	0.47	2.09	4.39	1.91	6.29	0.70
SuP14 <--- LR	0.44	0.19		0.56					
SuP26 <--- LR	0.54	0.29		0.46					
SuP62 <--- LR	0.58	0.34		0.42					
KB3 <--- LRe	0.54	0.29	0.35	0.46	2.34	5.49	1.66	7.15	0.77
KB27 <--- LRe	0.58	0.34		0.42					
KB39 <--- LRe	0.56	0.31		0.44					
KB63 <--- LRe	0.67	0.44		0.33					
BSP64 <--- SLT	0.64	0.42	0.34	0.36	2.92	8.53	2.08	10.61	0.80
BSP52 <--- SLT	0.54	0.30		0.46					
BSP28 <--- SLT	0.64	0.41		0.36					
BSP16 <--- SLT	0.60	0.36		0.40					
BSP4 <--- SLT	0.49	0.24		0.51					
PTP53 <--- AQT	0.69	0.48	0.42	0.31	2.57	6.62	1.43	8.05	0.82
PTP41 <--- AQT	0.67	0.45		0.33					
PTP17 <--- AQT	0.63	0.40		0.37					
PTP5 <--- AQT	0.58	0.34		0.42					
PHD9 <--- ASe	0.58	0.33	0.37	0.43	3.04	9.24	1.96	11.20	0.83
PHD21 <--- ASe	0.61	0.37		0.39					
PHD45 <--- ASe	0.65	0.42		0.35					
PHD57 <--- ASe	0.61	0.38		0.39					
PHD69 <--- ASe	0.59	0.35		0.41					
TM71 <---	0.70	0.49	0.33	0.30	2.25	5.06	1.75	6.81	0.74

AIn									
TM59 <--- AIn	0.53	0.28		0.47					
TM35 <--- AIn	0.61	0.38		0.39					
TM11 <--- AIn	0.41	0.16		0.60					
PA66 <--- AcI	0.52	0.27		0.48					
PA54 <--- AcI	0.65	0.42		0.35					
PA42 <--- AcI	0.66	0.43	0.31	0.34	2.73	7.45	2.27	9.72	0.77
PA18 <--- AcI	0.40	0.16		0.60					
PA6 <--- AcI	0.51	0.26		0.49					
MB68 <--- Fb	0.63	0.40		0.37					
MB56 <--- Fb	0.68	0.47		0.32					
MB44 <--- Fb	0.49	0.24	0.32	0.51	2.80	7.84	2.20	10.04	0.78
MB32 <--- Fb	0.51	0.26		0.49					
MB20 <--- Fb	0.49	0.24		0.52					
EBP10 <--- LEv	0.59	0.35		0.41					
EBP22 <--- LEv	0.60	0.37		0.40					
EBP46 <--- LEv	0.65	0.42	0.40	0.35	3.17	10.02	1.84	11.85	0.85
EBP58 <--- LEv	0.60	0.36		0.40					
EBP70 <--- LEv	0.72	0.52		0.28					
KRS67 <--- SPA	0.66	0.44		0.34					
KRS43 <--- SPA	0.68	0.46	0.36	0.32	2.32	5.40	1.68	7.08	0.76
KRS7 <--- SPA	0.62	0.39		0.38					
KRS31 <--- SPA	0.36	0.13		0.64					
KPB12 <--- QCL	0.68	0.46		0.32					
KPB24 <--- QCL	0.64	0.41		0.36					
KPB36 <--- QCL	0.62	0.39	0.37	0.38	3.64	13.24	2.36	15.60	0.85
KPB48 <--- QCL	0.65	0.43		0.35					
KPB60 <--- QCL	0.54	0.30		0.46					
KPB72 <--- QCL	0.50	0.25		0.50					

Discriminant Validity

Discriminant validity can be analysed using several methods that have been identified. Joreskog (1971) proposed that Discriminant Validity is determined by limiting the correlation between two constructs to a parameter of 1.0. Discriminant validity is achieved when the unconstrained model has a lower Chi-square value than the constrained model (Zakbar, 2000) and the correlation property is different from 1.00 (Schmitt and Stults, 1986). Then the difference of Chi-square analysis values for the restricted model with the unrestricted model is implemented. Another method is based on the proposal of Fornell and Lacker (1981), where the importance of variance shared between constructs should be calculated using correlation analysis between constructs. The correlations between the constructs should be squared and the data obtained compared with the AVE values. The squared correlation value must be smaller than the AVE value to support Discriminatory Validity. In addition, Discriminatory Validity can also be determined by comparing the value of Maximum Shared Variance (MSV) with the AVE. MSV values were obtained from analysing correlation values between two squared constructs. Next, all MSV values are averaged to get the Average Shared Squared Variance (ASV) deal. Discriminant validity will be achieved when both the MSV and ASV values obtained are smaller than the AVE values (Hair *et al*., 2010).

The method proposed by Fornell and Lacker (1981) and Hair *et al*. (2010) is based on the value of AVE as a determinant of achieving Discriminatory Validity. Nevertheless, the AVE values obtained for the 12 constructs studied were smaller than 0.50 and less suitable for using the proposed method. In addition, most recent studies have begun to question the practice concerning the sensitivity of the analysis in question in addressing the issue of Discriminatory Validity between constructs (Henseler *et al*., 2015). Thus, the method proposed by Joreskog (1971) was used in determining Discriminatory Validity. The findings show that the Chi-square value for the two pairs of variables whose correlation is limited by 1.00 is greater than the value for the unconstrained correlation. The result proves that all elements have Discriminatory Validity. The Discriminant Validity Analysis for the 12 constructs is as in Table 5.

Table 5 :Discriminant Validity

Pairwise	Constraint			Unconstraint		
	Chi-square	df:	p	Chi-square	df	p
LS <---> LR	289.45	20	0.00	130.78	19	0.00
LS <---> KB	342.15	20	0.00	161.39	19	0.00
LS <---> SLT	262.74	27	0.00	66.55	26	0.00
LS <---> AQT	274.69	20	0.00	46.76	19	0.00
LS <---> AcI	338.20	27	0.00	119.77	26	0.00
LS <---> SPA	256.69	20	0.00	112.09	19	0.00
LS <---> Fb	242.58	27	0.00	78.95	26	0.00
LS <---> ASe	215.70	20	0.00	60.97	19	0.00
LS <---> LEv	285.51	27	0.00	65.07	26	0.00
LS <---> AIn	307.99	20	0.00	112.52	19	0.00
LS <---> QCL	303.84	35	0.00	97.05	34	0.00
LR <---> LRe	317.27	20	0.00	60.70	19	0.00
LR <---> SLT	271.44	27	0.00	134.82	26	0.00
LR <---> AQT	282.83	20	0.00	59.08	19	0.00
LR <---> AcI	374.03	27	0.00	170.69	26	0.00
LR <---> SPA	249.68	20	0.00	108.99	19	0.00
LR <---> Fb	225.16	27	0.00	47.23	26	0.01
LR <---> ASe	207.03	20	0.00	81.38	19	0.00
LR <---> LEv	308.22	27	0.00	83.77	26	0.00
LR <---> AIn	278.61	20	0.00	62.74	19	0.00
LR <---> QCL	328.03	35	0.00	118.39	34	0.00
LRe <---> SLT	265.11	27	0.00	57.66	26	0.00
LRe <--- >AQT	269.76	20	0.00	54.20	19	0.00
LRe <---> AcI	310.83	27	0.00	99.58	26	0.00
LRe <---> SPA	273.56	20	0.00	134.06	19	0.00
LRe <---> Fb	248.06	27	0.00	83.22	26	0.00
LRe <---> ASe	183.91	20	0.00	39.33	19	0.00

LRe <---> LEv	219.76	27	0.00	53.21	26	0.00
LRe <---> AIn	258.51	20	0.00	73.23	19	0.00
LRe <---> QCL	242.92	35	0.00	72.52	34	0.00
SLT <---> AQT	204.49	27	0.00	49.27	26	0.00
SLT <---> AIn	224.45	35	0.00	78.95	34	0.00
SLT <---> SPA	146.95	27	0.00	88.19	26	0.00
SLT <---> Fb	158.36	35	0.00	58.32	34	0.00
SLT <---> ASe	333.09	27	0.00	241.35	26	0.00
SLT <---> LEv	241.90	35	0.00	78.81	34	0.00
ALT <---> AIn	296.92	27	0.00	170.97	26	0.00
ALT <---> QCL	221.67	44	0.00	72.01	43	0.00
AQT <---> AcI	291.04	27	0.00	101.34	26	0.00
AQT <---> SPA	194.80	20	0.00	72.48	19	0.00
AQT <---> Fb	201.04	27	0.00	51.79	26	0.00
AQT <---> ASe	185.99	20	0.00	38.83	19	0.01
AQT <---> LEv	141.52	27	0.30	29.06	26	0.31
AQT <---> AIn	226.17	20	0.00	39.27	19	0.00
AQT <---> QCL	237.24	35	0.00	52.13	34	0.02
AcI <---> SPA	323.52	27	0.00	229.57	26	0.00
AcI <---> Fb	220.87	35	0.00	115.00	34	0.00
AcI <---> ASe	230.29	27	0.00	133.75	26	0.00
AcI <---> LEv	238.86	35	0.00	76.80	34	0.00
AcI <---> AIn	248.55	27	0.00	115.35	26	0.00
AcI <---> QCL	282.78	44	0.00	140.50	43	0.00
SPA <---> Fb	348.94	27	0.00	252.20	26	0.00
SPA <---> ASe	298.93	20	0.00	180.96	19	0.00
SPA <---> LEv	411.52	27	0.00	220.56	26	0.00
SPA <---> AIn	305.20	20	0.00	158.56	19	0.00
SPA <---> QCL	282.58	35	0.00	102.72	34	0.00
Fb <---> ASe	223.96	27	0.00	105.86	26	0.00
Fb <---> LEv	286.02	35	0.00	99.87	34	0.00
Fb <---> AIn	244.55	27	0.00	95.21	26	0.00
Fb <---> QCL	240.73	44	0.00	70.80	43	0.01
ASe <---> LEv	267.05	27	0.00	67.75	26	0.00
ASe <---> AIn	260.17	20	0.00	100.22	19	0.00
ASe <---> QCL	252.29	35	0.00	76.56	34	0.00
LEv <---> AIn	367.91	27	0.00	197.42	26	0.00

LEv <---> QCL	289.66	44	0.00	104.95	43	0.00
AIn <---> QCL	202.65	35	0.00	75.05	34	0.00

Construct Validity and Reliability

Construct validity using CFA analysis should also be reported along with Composite Reliability (CR). Hair et al. (2010) stated that CR is an alternative to Cronbach's Alpha, specifically using Structural Equation Model (SEM) analysis. Thus, the Construct Validity and Reliability of AfLPMQ are as in Table 6.

Table 6:Construct Validity and Reliability

Construct / Item			**Standard Factor Loading**	**t-value**
Learning Standards (C.R. = 0.66)				
SP49	-	I design various assessment instruments to measure student progress in Mathematics (examples: quizzes, worksheets, mini-project assignments).	0.67	***
SP25	-	I plan the content of learning Mathematics from easy to difficult level.	0.55	8.18
SP1	-	I refer to the Curriculum and Assessment Standards Document (CASD) to plan Mathematics lessons.	0.40	6.49
SP13	-	I plan to learn according to the students' cognitive level, from low to high thinking.	0.36	6.02
Learning Resources (C.R. 0.70)				
SuP62	-	I asked students to show Maths answers using a mini whiteboard to track learning progress.	0.58	***
SuP26	-	I use technology platforms (example: Learning Management System/Digital Learning/Schoology) to share Mathematics learning materials.	0.54	8.57
SuP2	-	I use technological materials (example: video/image/text/audio/graphics) to explain the basic concepts of Mathematics.	0.53	8.15
SuP14	-	I give Math exercises using online applications (Example: Kahoot/quizizz/socrative).	0.44	7.05
Learning Readiness (C.R. 0.77)				
KB63	-	I designed an induction set involving digital animation/video projection/singing/puzzles to engage students in learning Mathematics.	0.67	***
KB27	-	I practice rotating the role of students as leaders/group members in the learning of Mathematics.	0.58	11.37
KB39	-	I grouped students in small groups consisting of various levels of ability during the learning of Mathematics.	0.56	10.66
KB3	-	I apply drill techniques as the primary method in teaching Mathematics.	0.54	9.76
Sharing Learning Targets and Success Criteria (C. R. = 0.80)				
BSP4	-	I use stimuli (example: video/image/slide) to stimulate Math problem-solving skills.	0.49	***
BSP28	-	I use thinking tools such as i-think maps/graphic management in stating Mathematics learning objectives.	0.64	9.91
BSP64	-	I stimulate students to achieve success/optimal score criteria by giving rewards (tokens/prizes/certificates).	0.64	9.84
BSP16	-	I discussed the achievement of the learning objectives of Mathematics.	0.60	9.74
BSP52	-	I displayed the learning objectives on the column and asked students to read aloud.	0.54	9.39
Application of Questioning Techniques (C.R. = 0.82)				
PTP5	-	I ask open-ended questions to stimulate high-level thinking (Example: Why/How).	0.58	***
PTP53	-	I use questioning about cause and effect to stimulate students to answer.	0.69	10.65
PTP41	-	I summarise students' answers in an easy-to-understand form (example: diagram/mind map/data representation/Mathematical sentence).	0.67	10.34
PTP17	-	I give time to think before the student answers the questions posed.	0.63	10.02

Active Involvement (C.R. = 0.77)				
PA66	-	I perform role-playing activities in teaching Mathematics by involving characters from various races in Malaysia (example: Malay/Chinese/Indian).	0.52	***
PA42	-	I prepare a group presentation schedule for Mathematics learning activities.	0.66	9.60
PA54	-	I asked students to express their emotions through activities (role play/acting/simulation/storytelling) during Mathematics learning by writing reflections in their exercise books.	0.65	9.48
PA6	-	I implement cross-curricular teaching in Mathematics learning (example: combining RBT/PSV subjects in a topic/space skills).	0.51	8.30
PA18	-	I apply Mathematics skills by linking them to solving routine problems/daily life of students.	0.40	6.97
Self and Peer Assessment (C.R. = 0.76)				
KRS7	-	I asked students to assess their understanding using traffic lights cards/emoticon cards.	0.62	***
KRS43	-	I encourage students to record new knowledge learned in the parking lot space.	0.68	12.05
KRS67	-	I give students space to express comments/suggestions on the work of other groups.	0.66	11.37
KRS31	-	I guide students to identify errors in Maths calculation steps.	0.36	6.77
Feedback (C.R. = 0.78)				
MB20	-	I apply hand gesture techniques (thumbs up/down) for students to agree with their friend's ideas/answers.	0.49	***
MB56	-	Student representatives from each group presented answers for review at the end of the activity period for feedback purposes.	0.68	9.25
MB68	-	I formulate student progress based on scores resulting from self and peer feedback.	0.63	9.02
MB32	-	I explain specifically the corrective steps that students need to take.	0.51	7.87
MB44	-	I show examples of answers to encourage students to improve the quality of assignments.	0.49	7.85
Appreciation and Self-esteem (C.R. = 0.83)				
PHD9	-	I give rewards (examples: stationery, candy) to students who answer questions correctly.	0.58	***
PHD45	-	I make thunderous applause for individual/group work a routine at each presentation session.	0.65	9.92
PHD21	-	I use the accumulated merit system in determining the best group for a Mathematics title to redeem reward points at the end of the period.	0.61	9.68
PHD69	-	I give recognition (example: certificate/button badge/name tag) to students who show progress.	0.59	9.34
PHD57	-	I spend time with students who are less proficient in learning Mathematics to review their problems.	0.61	9.22
Learning Evidences (C.R. = 0.85)				
EBP70	-	I formulate the level of understanding based on tools such as traffic lights/emoticon cards shown by students.	0.72	***
EBP46	-	I used checklists to make observations of group assignments where the information was analysed for improvement purposes.	0.65	12.06
EBP22	-	I recorded student responses in formative assessment as a reflection of teaching.	0.60	11.54
EBP58	-	I record the scores obtained for each Mathematics exercise (example: worksheets/online quizzes/projects).	0.60	11.42
EBP10	-	I recorded the progress of Math learning for each class taught.	0.59	11.11
Analysing Information (C.R. = 0.74)				
TM11	-	I summarise the number of students who have/have not mastered the skills taught while writing reflections.	0.41	***
TM71	-	I measure students' emotions in learning Mathematics using emoticon cards to obtain information on teaching performance.	0.70	8.04
TM35	-	I examine the way students think while solving Math problems as a diagnostic step.	0.61	7.93

TM59	-	I evaluate the effectiveness of high-level questions in stimulating student thinking.	0.53	7.36
Quality of Continuous Learning (C.R. = 0.85)				
KPB72	-	My committee members and I organised various Mathematics excellence programs at the school level (for example, Monthly Mathematics Literacy Quiz/Answering Techniques Workshop/Mathematics Project) to stimulate students' interest in the subject of Mathematics.	0.50	***
KPB12	-	I repeat the teaching of skills that students less master.	0.68	9.33
KPB48	-	I run a remedial program where high-achieving students help low-achieving students learn mathematics.	0.65	9.21
KPB24	-	I use other methods to teach skills that many students do not master.	0.64	8.94
KPB36	-	I teach students who do not master mathematics individually based on a specific topic/skill.	0.62	8.89
KPB60	-	I set Math learning goals between students and parents by writing learning contracts.	0.54	8.41

Model Fit Statistics (p = 0.00, RMSEA = 0.08, ratio = 3.39, PCFI = 0.66)
*** Item constrained for identification purposes
C.R. = Composite Reliability

DISCUSSION

CFA conducted through the first-order model analysis proved that AfLPMQ has construct validity when reaching a statistically defined model fit. However, each construct cannot achieve Convergent Validity due to the low AVE value obtained. This situation occurs due to the decision not to drop items with common loading factor values. At the same time, this poor-performing indicator can create a more unexplained variance in the model and ultimately hurt the ability to achieve Convergent and Discriminant Validity (Collier, 2020). According to Groves (1989), variance is a measure of variation in the data observed. It exists when a respondent does not answer a sensitive question honestly for various reasons, or the respondent may misinterpret or make a mistake in answering the question. These factors also increase the measurement error. In addition, the pilot study was conducted using the online survey method. At the same time, Cresswell (2012) suggested that the pilot study be conducted face to face with the respondents to get their feedback on the items developed.

However, this study has sought to minimise errors, as Groves (1989) stated. Among the errors of coverage that have been resolved involving all 91 primary schools in Hulu Langat, Selangor. Simple random sampling has been applied where five to six Mathematics teachers in each primary school have the opportunity to be selected to minimise sampling errors. Similarly, there is no missing value with nonresponse errors where each item is set as required to respond in the google form application. The high number of respondents measures the extent to which the study results can be generalised. However, measurement errors cannot be avoided entirely. In addition, the AfLPMQ may cause anxiety among respondents cause of no guidance. Therefore, it is recommended that face-to-face online discussions using various platforms such as google meet, zoom, webex, Microsoft team, and so on are necessary to implement a more efficient pilot study to obtain better statistical values. All the items should be discussed one by one to get feedback for improvement.

The decision to maintain the poor-performing AfLPMQ items was made after the AfL strategy theory obtained through a systematic review conducted during the conceptualisation phase. Therefore, stakeholders must re-examine the implementation of AfL in schools, especially for Mathematics subjects. The findings of this study are expected to help the Curriculum Development Division to refine the existing Classroom Assessment Implementation Guide, especially formative assessment that uses the AfL approach. The elements of AfLPMQ can be used to form a checklist for observation to improve teachers practised on AfL at the school administration level to ensure that AfL in TnL Mathematics is practised effectively. From the teacher's perspective, AfLPMQ can be used as a guide to ensure that the practices practised are on the right track so that students' potential can be developed comprehensively. Indirectly, the country's aspiration to produce holistic individuals is achievable.

CONCLUSION

The pilot study is the last step in the AfLPMQ development process. Decisions were made based on the theories underlying the conceptualisation of this study. The data obtained were analysed statistically using Confirmatory Factor Analysis to determine to Construct Validity and Composite Reliability. Although the tested AfLPMQs had achieved construct validity, there were still indicators that were not statistically achievable due to poor-performing items being maintained. Therefore, it is necessary for the implementation of AfL in TnL Mathematics in primary schools to be scrutinised so that the desired aspirations can be fully achieved.

REFERENCES

1. Akhmedina, A. (2017). Challenges of implementing formative assessment at Nazarbayev Intellectual School. MA thesis. Nazarbayev University Graduate School of Education.
2. Ali, E., & Kerpčarová, T. (2019). *Family Buying Behaviour : Parents ' perspective of children influence on their buying behaviour .* Master Thesis.
3. Anagün, Ş. S. (2018). *Teachers ' Perceptions about the Relationship between 21st Century Skills and Managing Constructivist Learning Environments. 11*(4), 825–840.
4. Arumugham, K.S (2020). Curriculum, Teaching and Assessment in The Perspective of Classroom Assessment. *Asian People Journal*, 3 (1) : 152-161.
5. Arumugham, K.S. (2018). Merancang, melaksana dan mentaksir dalam konteks pentaksiran berasaskan sekolah : Kajian kes guru-guru Prinsip Perakaunan Tingkatan Empat. Tesis Doktor Falsafah. Universiti Utara Malaysia.
6. Asamoah, D. (2019). Gender Difference in Formative Assessment Knowledge of Senior High School Teachers in the Upper West Region of Ghana. *Journal of Education and Practice*, (February), 8–13. https://doi.org/10.7176/jep/10-6-08
7. Bagozzi, R.P. Yi, Y., & Phillips, L.W. (1991). "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36(3), 421-458.
8. Bezabih, B. M., Yigzaw, Y. H., Ahmed, A., & Candidate, G. (2019). *Knowledge, Perception , Practices and Challenges of Assessment for Learning (AfL) among First Cycle Primary School Teachers in Somali Regional State. 24*(1), 21–33. https://doi.org/10.9790/0837-2401062133
9. Byrne, M.B. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming (Second Edition)*. New York : Routledge.
10. Chemeli, J. (2019). Impact of the five key formative assessment strategies on learner's achievement in mathematics instruction in secondary schools: A case of Nandi County, Kenya. International Academic Journal of Social Sciences and Education, 2(1), 212-229.
11. Cohen, L., Manion, L., and Morrison, K. (2018). *Research methods in educations.* New York: Routledge.
12. Collier, J. E. (2020). *Applied Structural Equation Modeling using AMOS : Basic to advanced techniques*. New York : Routledge.
13. Couper, Mick P. (2000) Review: web surveys: a review of issues and approaches. *The Public Opinion Quarterly*, 64 (4): 464–94.
14. Creswell, J.W. (2012). *Planning, Conducting, and Evaluating Quantitative and Qualitative Research (Fourth Edition)*. Boston : Pearson.
15. Crocker, L., Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. United States of America : Cengage Learning.
16. Danielson, C. (2017). *Assessment for learning — For teachers as well as students.* In The Future of Assessment (pp. 191-213). Routledge.
17. DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, *61*(March), 159–169. https://doi.org/10.1016/j.stueduc.2019.03.012
18. Duckor, B., & Holmberg, C. (2019). Exploring How to Model Formative Assessment Trajectories of Posing-Pausing-Probing Practices: Toward a Teacher Learning Progressions Framework for the Study of Novice Teachers. *Journal of Educational Measurement*, *56*(4), 836–890. https://doi.org/10.1111/jedm.12239
19. Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. Statistical methods for rates and proportions, 2(212-236), 22-23.
20. Fornell, Claes and David F. Larcker. (1981), "Evaluating Structural Equation Models and Unobservable Variables and Measurement Error", Journal of Marketing Research, 18 (February), 39–50.
21. Frey, B. D. (2018) *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. California : SAGE Publications
22. Gan, Z., He, J., & Mu, K. (2019). Development and Validation of the Assessment for Learning Experience Inventory (AFLEI) in Chinese Higher Education. *The Asia-Pacific Education Researcher*, *28*(5), 371–385. https://doi.org/10.1007/s40299-019-00435-7
23. Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley.

24. Guralnik, D. B. (Ed.). (1976). *Webster's new world dictionary of the American language (2nd College Ed.)*. Cleveland, OH: William Collins and World Publishing.
25. Hair, J.F. Jr., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate Data Analysis* : 7 Edition. Pearson.
26. Henseler, Jorg, Christian M. Ringle, and Mario Sarstedt. (2015), "A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling", *Journal of the Academy of Marketing Science*, 43 (1), 115–135.
27. Heredia, S. C., Furtak, E. M., Morrison, D., & Renga, I. P. (2016). Science Teachers' Representations of Classroom Practice in the Process of Formative Assessment Design. *Journal of Science Teacher Education*, *27*(7), 697–716. https://doi.org/10.1007/s10972-016-9482-3
28. Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: achievement, identity, and equity. *ZDM - Mathematics Education*, *50*(4), 729–741. https://doi.org/10.1007/s11858-018-0943-3
29. Hu, L. and P.M. Bentler. (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives", Structural Equation Modeling, 6 (1), 1–55.
30. Irwing, P., & Hughes, D. J. (2018). *Test Development*. The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development, First Edition.
31. Jamil, H & Said, R. R. (2019). Pelaksanaan penskoran pentaksiran lisan Bahasa Melayu dalam Pentaksiran Bilik Darjah. *Jurnal Pendidikan Bahasa Melayu*, Vol 9, Bil 2.
32. Joreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 109-133.
33. Kadir, A., Nor, M., & Hutagalung, F. D. (2019). Pemupukan Amalan Pendidikan Anti-Bias (Apab) Dalam Pendidikan Prasekolah Di Luar Bandar. Jurnal Kurikulum & Pengajaran Asia Pasifik, 7(4).
34. Kementerian Pendidikan Malaysia (2019). *Panduan Pelaksanaan Pentaksiran Bilik Darjah (Edisi Kedua)*. Putrajaya : Bahagian Pembangunan Kurikulum.
35. Kementerian Pendidikan Malaysia (2017). *Dasar Pendidikan Kebangsaan (Edisi Keempat)*. Putrajaya : Bahagian Perancangan dan Penyelidikan Dasar Pendidikan.
36. Kementerian Pendidikan Malaysia (2018). *Kurikulum Standard Sekolah Rendah : Dokumen Standard Kurikulum dan Pentaksiran Matematik Tahun Empat*. Putrajaya : Bahagian Pembangunan Kurikulum.
37. Kementerian Pendidikan Malaysia (2013). *Pelan Pembangunan Pendidikan Malaysia 2013-2025 (Pendidikan Prasekolah hingga Lepas Menengah)*.
38. Kenyon, B. J. (2019). *Teachers ' Formative Assessment Use to Check for Understanding and to Adjust Instruction*. Walden University. ProQuest Dissertations Publishing, 2019. 13427400.
39. Khan, F. N., & Hussain, S. (2019). Classroom Assessment , Literacy and Practices of Teacher Educators in Pakistan Classroom. *Global Social Sciences Review*, IV(IV).
40. Kline, T.J.B. (2005). *Psychological Testing: A Practical Approach to Design and Evaluation*. London : Sage Publications.
41. Koy, V., Yunibhand, J., Angsuroch, Y., & Turale, S. (2017). Development and Psychometric Testing of the Cambodian Nursing Care Quality Scale. *Pacific Rim International Journal of Nursing Research*, *21*(4), 267–279.
42. Krejcie, R. V., & Morgan, D. W. (1970). *Determining Sample Size for Research Activities. Educational and Psychological Measurement, 30(3), 607–610.* doi:10.1177/001316447003000308
43. Leavy, P. (2017). *Research Design : Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*. New York : The Guilford Press.
44. Lele, D. M., Setiawan, D., Righu, S. T., & Barat, K. S. (2018). Clinical Supervision Instrument Development f or Junior High School Teacher Based on Android. *Journal of Educational Research and Evaluation*, *7*(1), 94–100. https://doi.org/10.15294/jere.v7i1.19512.
45. Lyon, C. J., Nabors, O. L., & Wylie, C. E. (2019). Working toward integrated practice: Understanding the interaction among formative assessment strategies*. The Journal of Educational Research, 1–14.*
46. Miller, L. A., Lovler, R. L., & McIntire, S. A. (2015). *Foundations of psychological testing: a practical approach (4th ed.)*. California: SAGE Publications Inc.
47. Mohamed, S., Kamis, A., & Ali, N. (2016). Gauging the assessment literacy of Malaysia's home economics teachers: An empirical study. *Geografia: Malaysian Journal of Society & Space*, *12*(3), 130–138.

48. Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Kuram ve Uygulamada Egitim Bilimleri, 18*(1), 85–118. https://doi.org/10.12738/estp.2018.1.0216
49. Phakiti, A., De Costa, P., Plonsky, L., & Starfield, S. (2018). *The Palgrave Handbook of Applied Linguistics Research.*United Kingdom : Springer.
50. Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using the SPSS program*. 6th Edition. New York : Open University Press.
51. Peters, J. M. & Stout, D. L. (2006). *Methods for teaching elementary school science (5th ed.).* Columbus, OH: Pearson Publishing.
52. Rahimi, S. A., & Sahragard, R. (2019). *The Mismatch between EAP Teachers ' Beliefs and Classroom Practices toward Formative Assessment in the EFL Context of Hormozgan*.
53. Schmitt, N, & Stults, D. M. (1986*).* Methodology Review: Analysis of Multitrait-Multimethod Matrices. *Applied Psychological Measurement*, 10(1) : 1-22.
54. Sekaran, U., & Bougie, R. (2016). *Research Methods for Business: A Skill-Building Approach (seventh edition)*. United Kingdom : Wiley.
55. Tabachnick, B.G. & Fidell, L.S. (2013). *Using multivariate statistics (6th edn)*. Boston: Pearson Education.
56. Teo, P. (2019). Teaching for the 21st century: A case for dialogic pedagogy*. Learning, Culture and Social Interaction, 21, 170–178.* doi:10.1016/j.lcsi.2019.03.009
57. The Research Advisors (2006). Sample Size Table. http://research-advisors.com
58. Ulfvarson, J., Oxelmark, L., & Jirwe, M. (2018). Assessment in clinical education: A comparison between a generic instrument and a course-specific criterion-based instrument. *Nordic Journal of Nursing Research*, *38*(4), 205715851876840. https://doi.org/10.1177/2057158518768404.
59. Vatharaj, R., Abdullah, A.G.K. & Ismail, A. (2015) The effect of teacher autonomy on assessment practices among Malaysian Cluster School Teachers. *International Journal of Asian Social Science*, 5(1) : 31-36.
60. Widaman, K.F. (1985). Hierarchically Nested Covariance Structure Model for Multitrait-Multimethod Data. Applied Psychological Measurement, 9 (10): 1-26. https://doi.org/10.1177%2F014662168500900101
61. Yidana, M. B., & Partey., A. P. (2018). Economics Teachers' Conceptions of Classroom Assessment: A Study of Senior High Schools in the Central and Ashanti Regions of Ghana. *International Journal for Innovation Education and Research*, *6*(10), 153–174. https://doi.org/10.31686/ijier.vol6.iss10.1176
62. Zakbar, V. (2000). Some Methodological Issues with Structural Equation Model Application in Relationship Quality Context.