Jurnal Pendidikan Malaysia 45(2)(2020): 79-91
DOI: http://dx.doi.org/10.17576/JPEN-2020-45.02-08

The Development and Validation of Assessment Practices of Mathematical Thinking (APMT) Instrument

(Pembangunan dan Pengesahan Instrumen Amalan Pentaksiran Pemikiran Matematik (APMT)

SAOUD ALHUNAINI, KAMISAH OSMAN*, & NASER ABDURAB

ABSTRACT

Assessment practices of mathematical thinking are an important part of mathematics teachers' practices. This survey aimed to develop and validate an Assessment Practices of Mathematical Thinking (APMT) instrument. A total of 523 mathematics teachers from Omani public schools were randomly selected as samples. The initial scale consisted of 25 items. The scale was validated by experts in mathematics education and educational measurement and evaluation. Exploratory and confirmatory factor analysis was used to test the APMT model. Results show that the goodness-of-fit indices for the first and second order of the APMT model are placed within the acceptable criteria, and that the magnitudes of loadings for all items are statistically significant. The final version of the APMT scale contains five factors and 21 items. In conclusion, this study has developed an APMT instrument is acceptable psychometrically, and useful to be used for assessing teachers' assessment practices of mathematical thinking in Arabic context, especially in Oman. The instrument can also be used by teachers for the purpose of self-assessment of their assessment practices of mathematical thinking, which can show their strengths and weaknesses. This instrument is unique because it focuses on different types of assessment practices that can be implemented to evaluate and develop students' mathematical thinking within and out of the classroom.

Key Words: Mathematical thinking; assessment practices; diagnostic assessment; formative assessment; summative assessment; alternative assessment; electronic assessment; confirmatory factor analysis; exploratory factor analysis; mathematics teachers

ABSTRAK

Amalan pentaksiran pemikiran matematik adalah bahagian yang penting dalam amalan guru matematik. Kajian tinjauan ini bertujuan untuk membina dan mengesahkan instrumen Amalan Pentaksiran Pemikiran Matematik (APMT). Sebanyak 523 guru matematik dari sekolah-sekolah awam di Oman dipilih secara rawak sebagai sampel. Skala awal terdiri daripada 25 item. Skala ini disahkan oleh pakar dalam pendidikan matematik dan pengukuran dan penilaian pendidikan. Analisis faktor eksploratori dan pengesahan digunakan untuk menguji model APMT. Hasil menunjukkan bahawa indeks goodness-of-fit untuk susunan pertama dan kedua model APMT diletakkan dalam kriteria yang diterima, dan magnitud muatan untuk semua item adalah signifikan secara statistik. Versi akhir skala APMT mengandungi lima faktor dan 21 item. Kesimpulannya, kajian ini telah membina sebuah instrumen APMT yang dapat diterima secara psikometrik, dan berguna untuk digunakan untuk menilai amalan penilaian pemikiran pemikiran matematik guru dalam konteks Arab, terutamanya di Oman. Instrument ini juga boleh digunakan oleh guru untuk tujuan penilaian kendiri amalan penilaian mereka terhadap pemikiran matematik, yang dapat menunjukkan kekuatan dan kelemahan mereka. Instrumen ini unik kerana memberi tumpuan kepada pelbagai jenis amalan penilaian yang boleh dilaksanakan untuk menilai dan mengembangkan pemikiran matematik pelajar di dalam dan di luar kelas.

Kata Kunci: Pemikiran matematik; amalan penilaian; penilaian diagnostik; penilaian formatif; penilaian sumatif; penilaian alternatif; penilaian elektronik; analisis faktor pengesahan; analisis faktor penerokaan; guru matematik

INTRODUCTION

Several educational researchers have considered assessment and measurement aspect as very important for the educational science, especially for teaching and learning practices (Amua-sekyi 2016; Ghaicha 2016). Traditionally, assessment has been related to tests and exams to be used for making management decisions (Jakeman & Letcher 2003). However, in the mid-1980s, many new terms have been produced within assessment literature such as direct and indirect assessment, performance assessment and original assessment. For example, Turner's et al. (2016) qualitative research supplied sufficient evidences about different ways used by teachers to measure, describe and develop students' mathematical thinking. Assessment approaches, therefore, have been developed to be more practical and this led to the term "assessment practices" to describe teachers' repeated assessment process of students' learning (Segers et al. 1999).

Several studies around the world have examined teachers' assessment practices from different aspects. For instance, Abed and Awwad (2016), Genc (2005), and Wallace and White (2014) investigated mathematics teachers' assessment practices as a part of their teaching practices. They have studied the assessment of teaching and assessment of learning. The results indicated that the practices of assessment of learning appear more than the practices of assessment as learning in teachers' practices. Whereas, some researchers such as Alkharusi et al. (2012; 2014), Genc (2005), and Suurtamm et al. (2010) focused on the forms of mathematics teachers' assessment practices. They reported that mathematics teachers should use different forms of assessment such as observation, questioning, quizzes, self-assessment, and journals.

Mathematical thinking has received a great interest from mathematics education researchers. It has been linked to the process standards of the National Council of Teachers of Mathematics (NCTM 2000), which consist of problem-solving, reasoning and proof, representation, mathematical communications, and mathematics connections (Yong & Sam 2008). The assessment practices of mathematical thinking can give teachers an excellent chance to examine students' mathematical thinking and provide feedback that can help both students and teachers to develop student learning. Thus, assessment practices are considered as a powerful tool for enhancing students' mathematical thinking. Teachers need a rich amount of information about student mathematical thinking and learning to assess their progressions. The information can be collected through different types of assessment practices within and out of the classrooms (Acar-Erdol & Yıldızlı 2018; Siemon et al. 2017).

In addition, previous researchers found that there is a relationship between teacher mindfulness about mathematical thinking and their assessment practices of students' mathematics learning. For example, Gibney (2014) stated that assessment tasks need to better address teachers' awareness about mathematical thinking to improve students' mathematical thinking. Teachers can determine their students' mathematical knowledge base on the information that is obtained from analysing the solutions and answers of the activities. These information give the teacher an insightful view about their students' mathematical thinking (Turner et al. 2016).

There were different methods of classifying teachers' assessment practices of mathematics learning. From the literature, some researchers classified mathematics assessment practices into diagnostic, formative and summative assessment (Acar-Erdol & Yıldızlı 2018; Dandis 2013; Genc 2005). Some other researchers divided mathematics assessment into traditional and alternative assessments (Abed & Awwad 2016; Alkharusi et al. 2012, 2014; Ghaicha 2016) whereas electronic assessment practices have been added as one of the teachers' assessment practices (Stacey & Wiliam 2013; Zahner et al. 2012).

In addition to that, a few research have introduced evidences of effectiveness of new assessment models that can be used to compare between traditional and standardized evolutional tools such as tests, rubrics and learning tasks. The assessment models have introduced some frameworks that help to assess the development of students' learning of basic skills (Dandis 2013; Genc 2005). For example, Yong and Sam (2008) suggested a framework for evaluating students' mathematical thinking, which consists of metacognition rating scale, instrument of students' performance, mathematical dispositions rating scale, and scoring rubric for mathematical thinking. Whereas, Graf and Arieli-Attali (2015) designed a model for assessing the complex thinking in mathematics focusing on the formative assessment.

Numerous studies around the world have investigated the assessment practices of mathematics teachers from different features. For example, Abed and Awwad (2016), Genc (2005), and Wallace and White (2014) examined the assessment practices of mathematics teachers which were applied within their teaching practices. They have examined different kinds of assessment practices such as assessment of learning (summative assessment), assessment for learning (formative assessment), alternative assessment, diagnostic assessment and electronic assessment. The results inferred that teachers used the summative assessment practices more frequently than the formative assessment practices. On the other hand, some research such as Alkharusi et al. (2012; 2014), Genc (2005), and Suurtamm et al. (2010) studied the forms of assessment practices used by mathematics teachers. They stated that mathematics teacher applied different forms of traditional assessment such as exams, oral question and quizzes or alternative like observation, journals and self-assessments. These practices help teachers increase the chances for examining students' mathematical thinking. They also provided feedback that can be useful for both teachers and students to improve students' learning.

According to Sadik (2011), electronic assessments are one of the assessment practices that can address the digital tools to collect data about the students learning. Electronic assessments are nowadays essential because electronic devices are widely used everywhere in the world. Furthermore, a lot of countries and educational institutions apply electronic learning and assessments as they are vital instruments in their educational system. The electronic assessments are offered for students through different resources such as calculators, website, computers, electronic applications, and tablets (Kimmel 2011). Consequently, it is important that the electronic assessments to be included within the model of mathematics teachers' assessment practices of mathematical thinking.

Assessment practices of mathematical thinking (APMT) refer to all frequencies and goal-oriented events, with a focused function to assess the knowledge and abilities of students related to the mathematical thinking. Therefore, assessment practices include all known types of assessment (diagnostic, summative, formative, alternative and electronic assessment). These assessment practices are applied in order to provide evidence about students' previous knowledge and the current progress of learning (Ghaicha 2016; Wallace & White 2014). The scale of APMT focus on the teachers' assessment practices of mathematical thinking. Previous researches introduced instruments to assess mathematics teachers' assessment practices in general context of mathematics, such as Alkharusi et al. (2017), Dandis (2013) and Genc (2005). Furthermore, the instruments developed in previous studies contained some factors that are included in the instrument of the current research. For example, the instrument of. Schoenfeld (2015) developed an instrument to assess mathematics teachers' assessment practices cover the summative and formative assessment, while Dandis (2013) developed an instrument to measure assessment practices of mathematics teachers that contained diagnostic assessments, formative assessments and summative assessments. Meanwhile, other instrument divided the assessment practices into traditional and alternative assessment practices, such as Alkharusi et al. (2012; 2014). On the other hand, there were some instruments like the one developed by Abed and Awwad (2016) that linked the assessment practices of mathematics teachers with the tools used to assess students mathematics learning.

However, the previous instruments of mathematics assessment were prepared to assess the mathematics teachers' assessment practices of mathematics learning in general and did not focus on teachers' assessment practices of mathematical thinking. Through a review of the literature, it was found that there is no study done in the Arab world, especially in Oman that developed and studied the psychometrics of a scale related to teachers' Assessment Practices of Mathematical Thinking (APMT). Therefore, this study tried to confirm the assessment practices of mathematical thinking that contained the factors: diagnostic assessment, summative assessment, formative assessment, electronic assessment, and alternative assessment, such as in Figure 1.

This study thus aimed to develop a scale of APMT and to test:

1. The reliability of Structural Equation Modelling (SEM) of teachers' APMT scale
2. The validity of SEM of teachers' APMT scale
3. The G factor of SEM of teachers' APMT scale

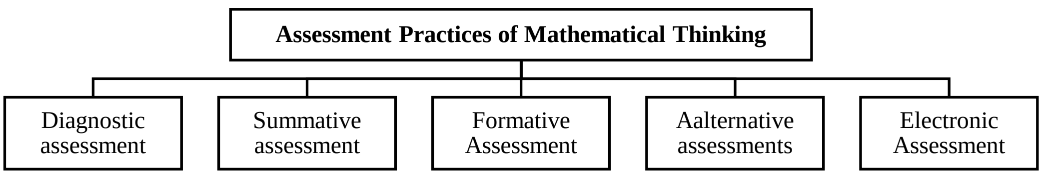

FIGURE 1. Factors of Assessment Practices of Mathematical Thinking

CONFIRMATORY FACTOR ANALYSIS (CFA) OF ASSESSMENT PRACTICES

Several scales of assessment practices have been tested by using confirmatory factor analysis. The researchers used several fit index tests such as standardized root mean square residual (SRMR), goodness-of-fit statistic (GFI), adjusted goodness-of-fit index (AGFI), incremental fit index (IFI), Tucker-Lewis index (TLI), comparative fit index (CFI), root mean square error of approximation (RMSEA) and ChiSq/df.

For example, Hasnida et al. (2018) tested a model of classroom assessment practices, which was a 5-points Likert scale instrument containing three factors: i) item administration and scoring, ii) preparation, and iii) utilization and evaluation. They administered the instrument to 320 teachers from Malaysian secondary schools. The reliability was tested using Cronbach's

Alpha for all factors specifically and comprehensively, and the values obtained were all more than 0.7. The results showed that all fit indices criteria were fulfilled (P-value=0.000, RMSEA=0.168, CFI=0.662, TLI=0.623, IFI=0.664 and ChiSq/df =10.033).

Ling et al. (2012) used an instrument of 5-point scale rating to assess the assessment practices of teachers from different subjects and levels. The instrument consisted of 5 dimensions. The five dimensions are constructing tests, types of assessment, use of assessment, grading and scoring and communicating assessment results. The CFA indices were NFI=0.924, CFI=0.928, GFI=0.955, AGFI=0.918, SRMR=0.041 and RMSEA=0.062.

Karaman and Sahin (2017) adapted a scale of formative assessment practices for Turkish culture. The scale was validated by using linguistic validity and applied to 400 teachers. The reliability of the scale was tested through Cronbach's Alpha reliability coefficient. Construct validity was found through finding the model fit (CFI=.90, RMSEA=.07, IFI=.90, TLI=.87).

METHODOLOGY

RESEARCH DESIGN AND INSTRUMENT

This research was conducted as a survey research on the first and second-order factor structure model, as done by Keith et al. (2006). This research has obtained ethical approval from the office of the Directorate General of Education North Sharqiyah Governorate, Ministry of Education of Oman.

The instrument used in this research was the APMT scale, which is a self-reported questionnaire that consists of 5-point Likert scale items as (1: Never; 2: Rarely; 3: Sometimes; 4: Mostly; 5: Always). The development of the scale in started by analysing a few scales of teachers' assessment practice that have been established in previous studies such as Abed and Awwad (2016), Alkharusi et al. (2012) and Genc (2005). Then, the items were checked if they can be used to measure teachers' assessment practices of mathematical thinking. Consequently, some items were chosen and linked with the standards of school mathematics (NCTM 2000), especially those that were more related to mathematical thinking skills. In other words, the items of the APMT scale were constructed by merging the NCTM standards and mathematics assessment practices (Diagnostic assessment, summative assessment, formative assessment, electronic assessment and alternative assessment). The items could be elicited from some items that have been used in preceding assessment practice scales. The items were written in Arabic language. The items were divided into 5 factors, as presented in Table 1.

TABLE 1. The number of items in each factor of APMT scale

APMT	Number of items
Diagnostic assessment	5
Formative assessment	5
Summative assessment	5
Alternative assessment	5
Electronic assessment	5
Total of items	25

Content validity of the APMT scale was achieved by presenting the items to a panel of experts in educational measurement and evaluation, mathematics education and educational psychology. They were asked to give their judgment about the clarity of wording, language and appropriateness of each item to be use by the school mathematics teachers in Oman, as well as the importance and comprehensiveness of the construct being measured. The feedbacks obtained from the experts were taken into account for editing the items. The suggested expert panels indicated that most items were related to the content of mathematics context in grades 5 through 12 and also related to the structures being evaluated. The judges indicated that the items were, in general, appropriately and clearly formulated to be understood by the Arabic mathematics teachers. However, they suggested some modifications in the construction of some items. For example, they changed the wording of some items to be clearer for the respondents to understand. The number of items in all factors have stayed the same.

The APMT scale reliability was obtained by calculating the Cronbach's Alpha coefficients for all factors of the APMT model. As presented in Table 2, all Cronbach's Alpha coefficient obtained for all factors are more than 0.7, and the overall Cronbach's Alpha coefficient for the instrument was 0.855, which indicated that all factors have met the required internal consistency.

TABLE 2. The Cronbach's Alpha coefficient for APMT factors

APMT factors	Cronbach's alpha coefficients
Diagnostic Assessment	.816
Formative Assessment	.816
Summative Assessment	.816
Alternative Assessment	.820
Electronic Assessment	.830
All items of APMT	0.855

SAMPLE

The population of the study consisted of all mathematics teachers who taught in the public schools in the Sultanate of Oman during the academic year of 2018/2019. From the population, a total number 537 mathematics teachers were sampled as the respondents in this study. They were randomly selected from the public schools in the educational governorates in the Sultanate of Oman. The samples have consented to this study and their identity and privacy were protected.

DATA ANALYSIS

The normality of items of each measurement scale was tested by using Skewness and Kurtosis. Factor should be placed within the satisfactory level, which is <7 for Kurtosis and <3 for Skewness (Kline 2015). Factor analysis was applied to examine the baseline constructs and issuing the construct validity for the assessment practices of mathematical thinking. The correlation matrix among the APMT scale items was explored at the start of any steps. The accepted value of each correlation should be <0.30 for orthogonal model, and ≥0.30 for oblique model (Tabachnick & Fidell 2007).

FINDINGS AND DISCUSSION

DESCRIPTIVE AND RELIABILITY ANALYSIS OF APMT INSTRUMENT

Table 3 presents the mean, standard deviation, skewness and kurtosis, and reliability of all dimensions and items for assessment practices of mathematical thinking. The results showed that mathematics teachers used diagnostic assessment to identify the student skills for learning the new topic in mathematics. They ask their students to work out tasks such as graphic representations to discover their ability of using the mathematical representations and to identify students' skills in mathematical thinking. Concerning the dimension of formative assessment, they employ oral questions and multiple-choice questions to train students on mathematical reasoning. They also provide their students with suggestions to test their ability of acquiring mathematical skills to develop their mathematical thinking skills.

Furthermore, mathematics teachers depend on the documents of assessment that are published by the Ministry of Education that describe and guide the assessment tools and the periods of applying them. They indicated that analysing students' results help them to determine the students' strengths and weaknesses in mathematical thinking skills. The teachers inform students about their performance at each assessment tool that has been used. They also consider the student's tendency towards mathematics when feedback is provided. Generally, the formative assessments were most frequently used assessment practices of the mathematics teachers, followed by diagnostic assessments, and the least used were alternative assessment.

In addition, the indicators of Skewness ≤ (-/+ 3) and Kurtosis ≤ (-/+ 7) suggest that all items of all factors are found within the standard threshold of Skewness ≤ (-/+ 3) and Kurtosis ≤ (-/+ 7). This shows that all items are normally distributed. Next, the Corrected Item-Total Correlation ≥ 0.30 for each item of all factors are positively related to the dimensions. The overall reliability for all items is .7 to .885, which are located within the accepted values (≤ 0.7) given by Kline (2015) as the conventional value for reliability.

TABLE 3. Mean, standard deviation, skewness and kurtosis, and reliability of dimensions of APMT

Items		M	SD	Rank	Skewness ≤ +/- 3	Kurtosis ≤ +/- 7	CITC ≥ 0.30
Diagnostic Assessment (DA)							
APMT1_D1	Apply diagnostic tests to identify students' skills in mathematical thinking.	3.69	.941	5	-.230	-.487	.424
APMT2_D2	Measure the level of students' mathematical thinking before starting to teach mathematics.	3.83	.868	3	-.542	.154	.525
APMT3_D3	Identify the skills that must have by students to learn a new topic in mathematics.	4.26	.776	1	-.761	-.125	.427
APMT4_D4	Ask students to study tasks such as graphic representations to discover their ability to represent the athlete.	3.82	.903	4	-.404	-.312	.503
APMT5_D5	Putting up mathematical problems to students before starting mathematics subjects.	3.93	.769	2	-.243	-.472	.416
	Overall Mean Score	3.91	0.851		Cronbach α =.702		
Formative Assessment (FA)							
APMT6_F1	Give students suggestions that enable them to monitor their progress in mathematical thinking skills.	3.95	.848	4	-.431	-.208	.463
APMT7_F2	Employ multiple choice questions to measure the level of students' ability of acquiring mathematical skills while my	4.15	.852	2	-.921	.806	.425

	mathematics teaching**.**						
APMT8_F3	Giving students pre-determined assignments and activities to develop their mathematical thinking abilities**.**	4.14	.802	3	-.628	.011	.505
APMT9_F4	Analyse students 'answers to identify students' progress in mathematical thinking skills.	3.88	.942	5	-.686	.194	.488
APMT10_F5	Employ oral questions to train students on mathematical reasoning.	4.15	.826	1	-.761		.323
	Overall Mean Score	4.05	0.854		Cronbach α =.702		
Summative assessment (SA)							
APMT11_S1	Consider the degree of student interest in learning when calculating the total degree in mathematics.	2.37	1.190	5	.753		-.267
APMT12_S2	Consider the extent of organisation of the home works and classroom activities when evaluating the level of mathematical thinking of the student**.**	3.78	.987	2	-.647		.148
APMT13_S3	Make a description of the mathematical thinking skills that each student has at the end of teaching mathematics course**.**	3.34	1.137	4	-.352		-.570
APMT14_S4	Inform each student of his or her strengths and weaknesses on the measuring instrument used to evaluate performance**.**	3.86	.919	1	-.450		-.358
APMT15_S5	Use Students' presentations to evaluate the level of mathematical thinking of them**.**	3.61	1.018		-.361		-.405
	Overall Mean Score	3.40	1.05		Cronbach α =.756		
Alternative Assessment (AA)							
APMT16_A1	Consider the student's tendency towards mathematics when feedback is provided.	3.80	.953	1	-.660		.237
APMT17_A2	Training students to evaluate strong and weak samples or models of classroom work related to the mathematical thinking of previous students**.**	3.40	1.168	3	-.361		-.723
APMT18_A3	Comparison of a student's level with the levels of other students in mathematical thinking.	3.20	1.248	4	-.264		-.963
APMT19_A4	Activating the student's portfolio to evaluate students' mathematical thinking.	3.62	1.110	2	-.567		-.339
APMT20_A5	Activation of computer programs to evaluate the level of mathematical thinking of the students**.**	3.15	1.207	5	-.106		-.902
	Overall Mean Score	3.43	1.137		Cronbach α =.771		
Electronic Assessment (EA)							
APMT21_E1	Assign students to perform some tasks using computerized mathematics programs in the implementation of graphs**.**	3.04	1.161	3	.103		-.777
APMT22_E2	Employing some phone applications for developing students' mathematical thinking skills**.**	3.04	1.102	2	-.105		-.548
APMT23_E3	Design of some electronic tests related to mathematical thinking skills.	2.91	1.232	5	.063		-.907
APMT24_E4	Activating websites related to mathematical thinking to train students on self-assessment.	2.92	1.214	4	.051		-.915
APMT25_E5	Use electronic calculators in training students on some mathematical conclusions.	3.27	1.141	1	-.289		-.669
	Overall Mean Score	3.04	1.17		Cronbach α =.855		

CONFIRMATORY FACTOR ANALYSIS OF APMT INSTRUMENT

APMT model was assessed through five components: Diagnostic Assessment (5 items), Formative Assessment (5 items), Summative Assessment (5 items), Alternative Assessment (5 items) and Electronic Assessment (5 items). As shown in Table 3, the goodness of fit indices for the first order CFA of the APMT model suggest that its fit statistics is less than the required criteria, as shown by the Normed Chi-Squared (CMINDF)=3.464 (not achieved the threshold of <3), Comparative Fit Index (CFI)=0.864 (less than the threshold of >0.90), Incremental Fit Index (IFI)=0.865 (not passed the threshold of >0.90), Tucker Lewis Index (TLI)=.847 (not passed the threshold of >0.90), and Goodness of Fit index (GFI)=.879 (not achieved the threshold of >0.90).

Besides, the magnitude of some items loadings is less than 0.50, which is an acceptable value at least. Subsequently, this baseline model needs to be improved until a plausible model is developed, which is done by removing items with loading below 0.50, as they less theoretically contribute to the shaping and modelling of their respective construct. Therefore, the third item from the diagnostic assessment dimension, the fifth item from the formative assessment dimension, the first item from the summative assessment dimension, and the fifth item from the alternative assessment dimension were deleted, as suggested by Modification Indices (MI) in AMOS.

As appeared in Figure 2 and Table 7, the goodness-of-fit indices for the first order of the respecified APMT model, show that its fit statistics are placed within the acceptable criteria, as shown by the Normed Chi-Squared (CMINDF)= 2.378 (achieved the threshold of <3), Comparative Fit Index (CFI)=0.934 (passed the threshold of >0.90), Incremental Fit Index (IFI)=0.935 (met the threshold of >0.90), Tucker Lewis Index (TLI)=0.923 (reached the threshold of >0.90), Goodness of Fit Index (GFI)=0.931 (accomplished the threshold of >0.90), Standardized Root Mean Residual (SRMR)=.043 (met the threshold of <0.80) and Root Mean Squared Error Approximation (RMSEA)=.051 (achieved the threshold of <0.80). These results inferred that the model is ready to be tested for the second order of APMT.

Furthermore, Table 4 shows that the magnitudes of loadings for all items are statistically significant (t-value≥1.964 and p-value≤0.05) with a value of 0.50 as acceptable coefficients for exploratory level. Composite Reliability (CR) for each construct in hypothesizing model met the acceptable criteria (0.70), while the Average Variance Extracted (AVE) did not meet the acceptable criteria (0.50). However, it can complete the further analysis as long as the Composite Reliability (CR) is achieved without meeting the AVE (Fornell & Larcker 1981a; 1981b), especially when the issues of construct validity (Convergent Validity and Discriminant Validity) for APMT model are beyond the objective of the current research.

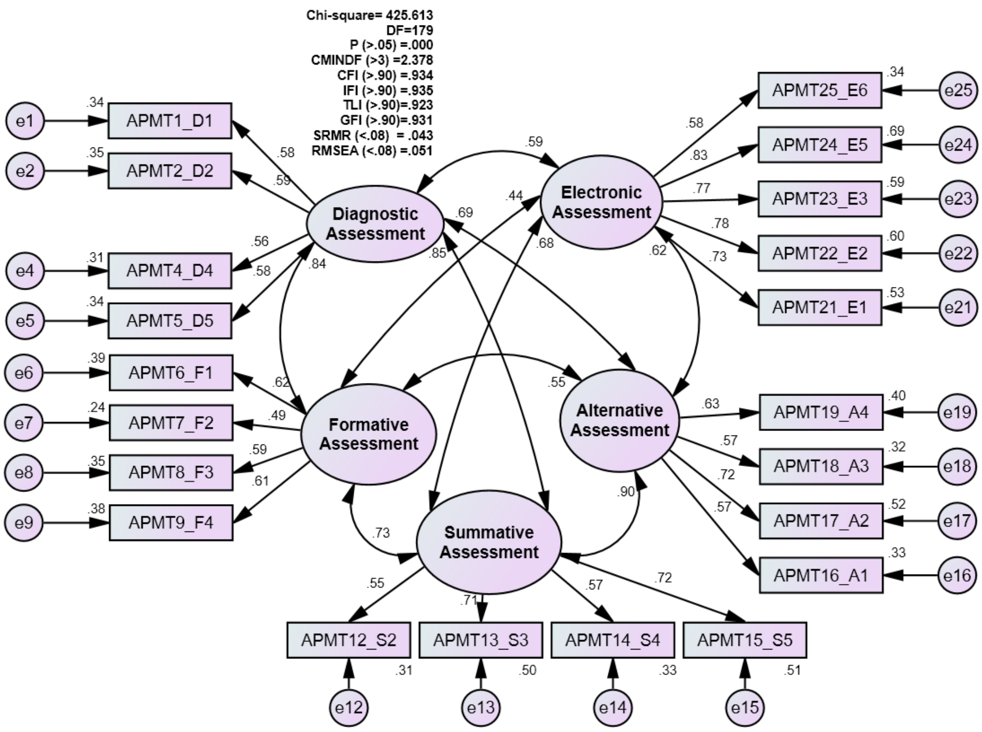

FIGURE 2. Confirmatory Factor Analysis for APMT (Respecified Model)

TABLE 4. Results of Model of APMT

Construct	Items	B	SE	T-Value	P	Factor loading	SMC	CR	AVE
Assessment Practices of Mathematical Thinking Component	Diagnostic Assessment	1.132	.103	11.004	***	.875	.766	0.925	0.713
	Formative Assessment	1.000				.778	.605		
	Summative Assessment	1.000				.995	.990		
	Alternative Assessment	1.372	.121	11.332	***	.863	.745		
	Electronic Assessment	1.002	.102	9.791	***	.677	.459		
Diagnostic Assessment	APMT1_D1	1.000				.580	.336	0.667	.333
	APMT2_D2	.936	.091	10.325	***	.589	.346		
	APMT4_D4	.920	.093	9.913	***	.556	.309		
	APMT5_D5	.823	.080	10.266	***	.584	.341		
Formative Assessment	APMT6_F1	1.000				.621	.386	0.669	.338
	APMT7_F2	.796	.089	8.952	***	.492	.242		
	APMT8_F3	.897	.087	10.249	***	.589	.347		
	APMT9_F4	1.097	.104	10.536	***	.614	.377		
Summative Assessment3	APMT12_S2	1.000				.555	.308	0.735	.413
	APMT13_S3	1.470	.124	11.845	***	.708	.502		
	APMT14_S4	.962	.093	10.339	***	.573	.328		
	APMT15_S5	1.331	.112	11.918	***	.716	.512		
Alternative Assessment4	APMT16_A1	.547	.041	13.181	***	.574	.330	0.718	0.391
	APMT17_A2	.837	.048	17.318	***	.718	.515		
	APMT18_A3	.706	.054	12.970	***	.566	.321		
	APMT19_A4	.700	.047	14.776	***	.631	.398		
Electronic Assessment5	APMT21_E1	1.271	.099	12.797	***	.730	.533	0.858	.551
	APMT22_E2	1.283	.097	13.290	***	.776	.602		
	APMT23_E3	1.422	.108	13.222	***	.769	.592		
	APMT24_E5	1.509	.109	13.780	***	.828	.686		
	APMT25_E6	1.000				.584	.342		

Key: B= Unstandardized Estimation, SE=Standard Error, P=Probability Value, SMC =Squared Multiple Regression, CR= Composite Reliability, AVE = Average Variance Extracted

TABLE 5. Results of Covariance among the constructs of APMT Model

Construct	Construct	B	SE	t-Value	p	r	SMC
Alternative Assessment	Summative Assessment	.494	.041	12.011	***	.903	0.81
Formative Assessment	Summative Assessment	.211	.027	7.912	***	.733	0.53
Electronic Assessment	Alternative Assessment	.414	.041	9.985	***	.621	0.38
Electronic Assessment	Formative Assessment	.154	.024	6.294	***	.439	0.19
Diagnostic Assessment	Alternative Assessment	.376	.038	9.907	***	.690	0.47
Diagnostic Assessment	Formative Assessment	.240	.029	8.328	***	.837	0.70
Formative Assessment	Alternative Assessment	.292	.034	8.471	***	.555	0.30
Electronic Assessment4	Diagnostic Assessment	.214	.029	7.352	***	.591	0.34
Electronic Assessment	Summative Assessment	.246	.031	7.864	***	.675	0.45
Diagnostic Assessment	Summative Assessment	.255	.031	8.236	***	.854	0.72

Key: B= Unstandardized Estimation, SE=Standard Error, P=Probability Value, r=Correlation, SMC =Squared Multiple Regression

Table 5 indicated that the results of testing covariances among the five constructs of the APMT model are statistically significant (t-value≥1.964 and p-value≤0.05). These results show that the correlations between each two factors are statistically significant.

As displayed in Table 6, the Square Root of Average Variance Extracted (AVE) for each construct of the hypothesized model was more than the Shared Variance (SV) (Multiple Squared Correlation (SMC)) with the exception for a few relationships, establishing the Discriminant Validity for the APMT model.

As depicted in Figure 3 and Table 7, the goodness-of-fit indices for the second order of CFA of the APMT demonstrate that its fit statistics are situated within the acceptable criteria, as shown by the Normed Chi-Squared CMINDF=2.594 (achieved the threshold of <3), Comparative Fit Index (CFI)=0.922 (passed the threshold of >0.90), Incremental Fit Index (IFI)=0.922 (met the threshold of >0.90), Tucker Lewis Index (TLI)=0.911 (reached the threshold of >0.90), Goodness of Fit Index (GFI)=.922 (accomplished the threshold of >0.90), Standardized Root Mean Residual (SRMR)=0.049 (met the threshold of <0.80) and Root Mean Squared Error Approximation (RMSEA)=.055 (achieved the threshold of <0.80). Finally, the magnitudes of higher loadings for five sub-factors are statistically significant (T-value ≥1.964 and P-value ≤0.05).

TABLE 6. Results of Shared Variance (SV) (Squared Multiple Correlation (SMC) and Square Root of AVE

Construct	1	2	3	4	5
Alternative Assessment	**0.53**				
Summative Assessment	0.81	**0.64**			
Formative Assessment	0.30	0.53	**0.58**		
Electronic Assessment	0.38	0.45	0.19	**0.74**	
Diagnostic Assessment	0.47	0.72	0.70	0.34	**0.58**

Bold size = Value of Squared Average Variance Extracted

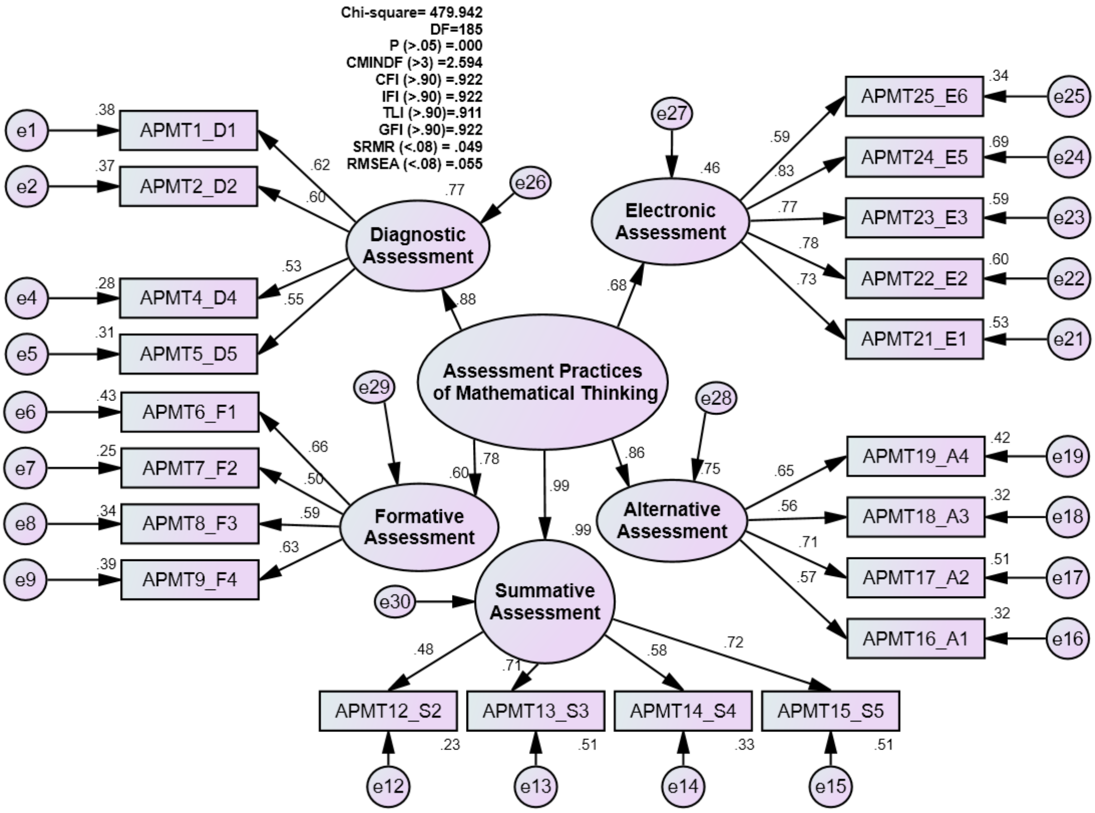

FIGURE 3. Second order of confirmatory factor analysis for assessment practices of mathematical thinking (Re-specified model)

TABLE 7. Goodness Fit Indices for Assessment Practices of Mathematical Thinking

	Indices	Required Scores	Baseline Model	Re-Specified Model	Second Order CFA
1	Chi-Square	-	918.090	425.613	479.942
2	DF (Degree of Freedom)	-	265	179	185
3	P (Probability)	(>.05)	.000	.000	.000
4	CMINDF (Normed Chi-Square)	(<3)	3.464	2.378	2.594
5	CFI (Comparative Fit Index)	(>.90)	.864	.934	.922
6	IFI (Incremental Fit Index)	(>.90)	.865	.935	.922
7	TLI (Tucker Lewis Index)	(>.90)	.847	.923	.911
8	GFI (Goodness of Fit Index)	(>.90)	.879	.931	.922
9	SRMR (Squared Root Mean Residual)	(<.08)	.056	.043	.049
10	RMSEA (Root Mean Squared Error Approximation)	(<.08)	.068	.051	.055

The final version of the APMT scale contains five factors and 21 items (Diagnostic assessment containing 4 items, formative assessment containing 4 items, summative assessment containing 4 items, alternative assessment containing 4 items and electronic assessment containing 5 items). The items were developed to collect enough information about the mathematics teachers' assessment practices of mathematical thinking. The assessment practices cover teachers' practices win and out of the classroom.

DISCUSSION

From the analyses shown, all factors have met the accepted level of reliability. Moreover, the result showed that the APMT model had a good fit of CFA. These results were consistent with several studies aimed to construct scales of teachers' assessment practices by using CFA, such as Hasnida et al. (2018), Karaman and Şahin (2017) and Ling et al. (2012).

The results illustrated teachers' practices of mathematical thinking in different situations and assessment type starting from the diagnostic assessment, passing through formative assessment and ending by summative assessment. The instrument also measured teachers' assessment practices related to alternative and electronic assessment, which are required nowadays to adapt teachers' practice to the new issues of educational practices and technology.

In addition, the results also indicate that teachers use diagnostic assessment to identify students' skills that are required for learning a new topic in mathematics. Generally, diagnostic assessment is used to define the difficulties that students face when they study new topics. It helps to fill the gap between the difficulties of what has been studied and what needed to be studied (Wallace & White, 2014).

The teachers use information from summative assessment to make a decision about the progress of their daily teaching. The results indicated that summative assessment involves informing each student of his or her strengths and weaknesses on the measuring instrument used to evaluate performance. Teachers used this information to decide whether they can move to new topics or to do more reviews. The primary purpose of summative assessment is to report valid and objective information about students' achievement at the end of the studying periods (Amua-sekyi 2016; Wallace & White 2014). The results are consistent with Wallace and White (2014), who demonstrated that teachers use summative assessment to collect information about the progress and understanding of students in response to their teaching.

Teachers benefit from their observation about students' reactions to their explanations and questions, especially oral questions to assess the students' interests of learning mathematics. Students' interest in learning is most often expressed through their motivation to perform academic tasks, their speed of completion and the quality of their output (Belbase 2015). This was supported by the finding which reported that the teachers consider student's interests towards mathematics based on the feedback provided. The teachers managed the amount of feedback and its appropriateness for their abilities. These results are in line with a study by Bremmer (2014), which showed that teachers manage the feedback given to the students. They considered the students' abilities when they give feedback, so as to tailor to their needs. Furthermore, previous studies also referred to the importance of appropriateness of feedback to the students' tendency and their abilities. Koloi-Keaikitse (2012) and Muñoz et al. (2012) emphasised that teachers should plan for the feedback that will be given to students. Feedback should be adapted to suit the students' situations for it to be useful for improving their learning. The results are consistent with Ling et al. (2012), who reported that teachers give their students' suitable feedback to their level of achievement.

Nowadays, the new mathematics curriculums pay more attention to advance mathematical thinking such as mathematical communications, connections, and reasoning skills (NCTM, 2000). Therefore, it is expected that the educational assessment system influences mathematics teachers in Oman. Mainly, they are asked to follow the documents of mathematics assessment that were prepared by the mathematics assessment team from the Ministry of Education in Oman. These documents contain multiple references to mathematical thinking skills like problem-solving, mathematical communications, and reasoning. The documents are used as the guideline for teachers to address the assessment tools and methods. The documents also defined the main concepts of assessment that helps to narrow the gap in understanding the assessment of teachers, so that the application is expected minimise them. However, the documents give teachers more freedom to apply the formative and alternative assessment with some suggestions that may help them. For example, the documents included some instructions that help teachers in assessing the oral assessment and how to assess the students' projects (Ministry of Education 2019).

This research has been unique from similar previous research done, that it presented an instrument for measuring mathematics teachers' assessment practices of mathematical thinking (APMT) that includes the dimensions of diagnostic assessments, summative assessment, formative assessment, alternative assessment and electronic assessment.

CONCLUSION

This study has tested the validity and reliability of the assessment practices of mathematical thinking (APMT) instrument. Results show that the goodness-of-fit indices for the first and second order of the APMT model are placed within the acceptable criteria, and that the magnitudes of loadings for all items are statistically significant. The final version of the APMT scale contains five factors and 21 items (Diagnostic assessment dimension containing 4 items, formative assessment dimension containing 4 items, summative assessment dimension containing 4 items, alternative assessment dimension containing 4 items, and electronic assessment dimension containing 5 items). In conclusion, this APMT instrument is acceptable and useful to be used for assessing teachers' assessment practices of mathematical thinking in Arabic context, especially in Oman. The scale can also be used by teachers for the purpose of self-assessment of their assessment practices of mathematical thinking, which can show their strengths and weaknesses. This instrument is unique because it focuses on different types of assessment practices that can be implemented to evaluate and develop students' mathematical thinking within and out of the classroom. The scale can be retested in the future with a wider range of sample and in another Arab country. This could reach to the improvement of the scale, which would be used for assessing the teachers' assessment practice of mathematical thinking.

REFERENCES

Abed, E. R., & Awwad, F. M. A. (2016). Students' Learning Assessment Practices Used by Jordanian Teachers of Mathematics for Grades (1-6). *International Education Studies*, *9*(1), 63–78.

Acar-Erdol, T., & Yıldızlı, H. (2018). Classroom Assessment Practices of Teachers in Turkey. *International Journal of Instruction*, *11*(3), 587–602.

Al-Musawi, A. (2009). Faculty Perceptions of the Professional Development Workshops Conducted at Sultan Qaboos University. *Journal of University Teaching & Learning Practice*, *5*(2), 92–104.

Alkharusi, H., Aldhafri, S., Al-Hosni, K., Al-Busaidi, S., Al-Kharusi, B., Ambusaidi, A., & Alrajhi, M. (2017). Development and validation of a scale for measuring mathematics teaching self-efficacy for teachers in the Sultanate of Oman. *International Journal of Instruction*, *10*(3), 143–158.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessemnt attitudes, competence, knowledge, and practices: An exploratory stud yof Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, *1*(2), 217–232.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Educational assessment profile of teachers in the Sultanate of Oman. *International Education Studies*, *7*(5), 116–137.

Amua-sekyi, E. T. (2016). Assessment , Student Learning and Classroom Practice : A Review. *Journal of Education and Practice*, *7*(21), 1–6.

Belbase, S. (2015). *Preservice secondary mathematics teachers' beliefs about teaching geometric transformations using geometric's sketchpad*. University of Wyoming.

Bremmer, A. L. (2014). *Teachers' knowledge of formative assessment initial instrument validation study*. Boise State University.

Brown, R., & Redmond, T. (2008). ReconceptuaTLIzing agency through teachers talking about a sociocultural approach to teaching mathematics in the classroom. In GoosMerrilyn, R. Brown, & K. Makar (Eds.), *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 101–108).Mathematics Education Reserach Group of Australasia Incorporated Prepared.Dandis, M. A. I. (2013). *Teachers' beliefs about assessment for learning: Introducing rubric in Secondary Education*. Universidad de Granada.

Fornell, C., & Larcker, D. F. (1981a). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Fornell, C., & Larcker, D. F. (1981b). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 382–388.

Genc, E. (2005). Development and validation of an instrument to evaluate science teachers' assessment beliefs and practices, Phd doctoral dessertation,Florida State Univercity.

Ghaicha, A. (2016). Theoretical Framework for Educational Assessment : A Synoptic Review. *Journal of Education and Practice*, *7*(24), 212–231.

Gibney, J. (2014). Provoking mathematical thinking: Experiences of doing realistic mathematics tasks with adult numeracy teachers. *Adults Learning Mathematics*, *9*(2), 97–115

Graf, E. & Arieli-Attali, M. (2015). Designing and developing assessments of complex thinking in mathematics for the middle grades. *Theory Into Practice,* 54(3), 195-202

Hasnida, N., Ghazali, C., Rabi, N. M., & Hassan, N. (2018). A Confirmatory Factor Analysis of Classroom Assessment Practises scale in a Malaysian Context. *International Journal of Academic Research in Progressive Education and Development*, *7*(3), 516–529.

Jakeman, A. J., & Letcher, R. A. (2003). Integrated assessment and modelling: Features, principles and examples for catchment management. *Environmental Modelling and Software*, *18*(6), 491–501.

Karaman, P., & Şahin, Ç. (2017). Adaptation of teachers' conceptions and practices of formative assessment scale into Turkish culture and a structural equation modeling. *International Electronic Journal of Elementary Education*, *10*(2), 185–194.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher Order , Multisample , Confirmatory Factor Analysis of the Wechsler Intelligence Scale for Children — Fourth Edition : What Does It Measure ? *School Psychology Review*, *35*(1), 108–127.

Kimmel, B. S. C. (2011). The school library: A space for mathematical thinking, learning, and sharing. *Library Media Connection*, *30*(3), 26–28.

Koloi-Keaikitse, S. (2012). *Classroom Assessment Practices: a Survey of Botswana Primary and Secondary School Teachers*. Phd doctoral dessertation, Ball State Universty.

Ling, S. S., Lan, O. L., Suah, S. L., & Ong, S. L. (2012). Investigating Assessment Practices of In-service Teachers. *International Online Journal of Educational Science*, *4*(1), 91–106.

Ministry of Education. (2019). *Document for assessing students' learning mathematics (5-8)*. Ministry of Education.

Muñoz, A. P., Marcela, P., & Escobar, L. (2012). Profile issues in teachers ` professional development teachers ' beliefs about assessment in an EFL context in Colombia. *PROFILE*, *14*(1), 143–158.

NCTM. (2000). *Principles Standards and for School Mathematics*. The National Council of Teachers of Mathematics, Inc.

Privitera, G. J. (2014). Survey and correlational research designs. In *Research Methods for the Behavioral Sciences* (pp. 227–240).

Sadik, A. M. (2011). A standards-based grading and reporting tool for faculty: Design and implications. *Journal of Education Technology*, *8*(1), 46–63.

Schoenfeld, A. H. (2015). Summative and Formative Assessments in Mathematics Supporting the Goals of the Common Core Standards. *Theory into Practice*, *54*, 183–194.

Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, *2*(2), 191–213.

Siemon, D., Day, L., Stephens, M., Horne, M., Callingham, R., Watson, J., Seah, R., & Siemon, D. (2017). Reframing Mathematical Futures : Using Learning Progressions to Support Mathematical Thinking in the Middle Years Developing Learning Progressions to Support Mathematical Reasoning in the Middle Years : Introducing the Reframing Mathematical Futures II Pr. In A. Downton, S. Livy, & J. Hall (Eds.), *Proceedings of the 40th Annual Conference of the Mathematics Education Research Group of Australasia* (p. 646).

Stacey, K., & Wiliam, D. (2013). Technology and assessment in mathematics. In *Third International Handbook of Mathematics Education* (pp. 721–751). Springer.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy and Practice*, *17*(4), 399–417.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). New Jersey: Pearson Education, Inc.

Turner, E. E., Foote, M. Q., Stoehr, K. J., McDuffie, A. R., Aguirre, J. M., Bartell, T. G., & Drake, C. (2016). Learning to Leverage Children's Multiple Mathematical Knowledge Bases in Mathematics Instruction. *Journal of Urban Mathematics Education*, *9*(1), 48–78.

Wallace, M., & White, T. (2014). Secondary Mathematics Preservice Teachers ' Assessment Perspectives and Practices : An Evolutionary Portrait 1. *Mathematics Education Research Group of Australasia*, *16*(2), 25–45.

Yong, H. T., & Sam, L. C. (2008). Implementing school-based assessment : The mathematical thinking assessment (MATA). *Buku Koleksi Bahan Seminar Inovasi Pedagogi IPBL Tahun 2008*, 73–88.

Zahner, W., Velazquez, G., Moschkovich, J., Vahey, P., & Lara-Meloy, T. (2012). Mathematics teaching practices with technology that support conceptual understanding for Latino/a students. *Journal of Mathematical Behavior*, *31*(4), 431–446.

Saoud Alhunaini
Faculty of Education
Universiti Kebangsaan Malaysia
Email: saod.alhunaini@moe.om

Kamisah Osman
Faculty of Education
University Kebangsaan Malaysia
Email: kamisah@ukm.edu.my

Naser Abdurab
Faculty of Education
Taiz University, Yeman
Email: nalareqe@yahoo.com

*Author for correspondence, email: kamisah@ukm.edu.my

Submitted: 18 August 2020
Reviewed: 13 October 2020
Accepted: 27 November 2020
Published: 30 November 2020