ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 13, No. 1, pp. 38-45, January 2022
DOI: https://doi.org/10.17507/jltr.1301.05

Establishing the Validity and Reliability of the Malaysian English Language Textbook Evaluation Checklist (MELTEC) Using Rasch Measurement Model (RRM)

Faustina Roberts
Faculty of Education, National University of Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

Azlina Abdul Aziz
Faculty of Education, National University of Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

Mohd Effendi @ Ewan Mohd Matore
Faculty of Education, National University of Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

***Abstract*—A textbook is an important teaching and learning material. Evaluating a textbook is essential to ensure educational goals are met. In the present study, the Malaysian English Language Textbook Evaluation Checklist (MELTEC) is developed for the purpose of materials adaptation. The lack of validity and reliability of previous textbook evaluation checklists reported by researchers in the field of ELT textbook evaluation has resulted in the need to establish the validity and reliability of the checklist developed in this study using the Rasch Measurement Model (RRM). One hundred and ten ESL secondary school teachers from Kuala Lumpur, Malaysia, had participated in this study. Rasch analysis was used to examine the item fit, polarity, reliability, separation index, and unidimensionality of the checklist. As a result, the finalised version of MELTEC comprises of 78 items. A total of 63 items were dropped as these items were not deemed to be suitable. The analysis performed using Rasch Measurement Model proved that the checklist has high validity and reliability. The checklist is found to be suitable to evaluate ELT secondary school textbooks in Malaysia. On that account, future researchers can consider using Rasch Measurement Model to establish the validity and reliability of the instruments developed to ensure their integrity and quality.**

***Index Terms*—validity, reliability, textbook evaluation checklist, Rasch Measurement Model (RRM)**

I. Introduction

A textbook will never be suitable for all teaching and learning situations despite how well it is written (Nguyen, 2015). Hence, textbook evaluation is pivotal to ensure educational goals are met. There are various methods to evaluate an ELT textbook, and the checklist method is one of the methods commonly employed. There are a number of advantages of using the checklist method. The most prominent advantage would be that it is economical and systematic (Cunningsworth, 1995; McGrath, 2002). Besides, the explicit criteria in a checklist allow for a thorough evaluation while reducing impressionistic judgment (Nimehchisalem & Mukundan, 2013). Moreover, evaluators can also add, modify and adapt the criteria within a checklist to suit the context of evaluation (Mukundan & Ahour, 2010; Demir & Ertas, 2014). The ability to transform a textbook evaluation checklist to make it suitable for the context of evaluation, makes this method of evaluation unique. One of the purposes of a textbook evaluation checklist is to facilitate ESL teachers to make informed decisions on materials adaptation. Adaptation is when ESL teachers make partial or minor changes in a textbook in an attempt to make it suitable for a particular class or level (Nehal, 2016). Adaptations are necessary despite careful selection of textbooks (Halim & Halim, 2017; McDonough, Shaw & Masuhara, 2013). Despite how suitable learning materials are, they may not cater to the different needs, learning styles, and cultural norms of individual learners (Tomlinson, 2012).

Numerous textbook evaluation checklists have been developed by previous researchers, albeit these checklists may not be suitable for the Malaysian context. Evaluation criteria are best customised to suit a specific context to appeal for local use (Isik, 2018). Moreover, textbook evaluation checklists in the past suffer from context-sensitivity (Simsek & Dündar, 2018). These checklists used items in their predecessors or were reworded or reclassified to a more practical version. They may not reflect the actual picture of a textbook and meet the preference of the end-users. Nimehchisalem and Mukundan (2015) also argue that checklists in the literature lack validity and reliability. On record, there are also only a few studies that test the efficiency of the checklists developed (Mukundan & Ahour, 2010; Isik, 2018). Therefore, the Malaysian English Language Textbook Evaluation Checklist (MELTEC) was developed in the present study to

evaluate the ELT textbooks used for the teaching and learning of English in secondary schools. The checklist is to be used by ESL teachers in the secondary schools in Malaysia for materials adaptation. As aforementioned, the literature highlights the lack of validity and reliability of the textbook evaluation checklists developed in the past. On that account, to establish the reliability and validity of MELTEC, a series of methods are employed, among which is the Rasch Measurement Model. The Rasch Measurement Model, a one-parameter Item Response Theory (IRT), is a modern model of measurement often used in the social sciences. The weaknesses of methods of analysis that fall under the Critical Test Theory (CTT), such as the Exploratory factor analysis (EFA) and Confirmatory factor analysis (CFA) in determining validity and reliability, lead to the reason the Rasch Measurement Model was chosen in the present study.

The Rasch Measurement Model has been the center of attention by various researchers worldwide in which their interest lies in building and constructing new instruments. The model by Georg Rasch is often a preferred choice among researchers due to the advantages that constructs, such as linearity, independence, objectivity, comprehensiveness, and inferences, can be easily deduced (Wright & Stone, 1979). This model advocates that response towards an item is solely influenced by an individual's ability and item difficulty (Bond & Fox, 2015). Rasch model ascertains that each item is only constructed based on the parameter of difficulty. An individual may have a 50:50 chance of attempting the items correctly given the difficulty logits of 0.00. Consequently, it is to be accepted that when the item difficulty is increased, the chances of success will be affected and in this case decreased. The order of evaluation within the model may vary according to the needs of a particular study. Under the Rasch Measurement Model, there are eight diagnostic data analyses involved in the process of instrument development, which include (i) unidimensional; (ii) compatibility (fit) item; (iii) polarities item; (iv) reliability and separation item respondents; (v) appropriateness of the measurement scale based on the use of categories; (vi) value of standardised residual correlation in determining leaning item; (vii) differential items functioning (DIF) based on gender; and (viii) the distribution of item difficulty levels and abilities of respondents (Hassan, 2012). However, based on the objective and needs of the present study, the following areas of analysis are performed using the Rasch Measurement Model; (i) item fit; (ii) item polarity (iii) reliability and separation index and (iv) unidimensionality. These analyses are sufficient to establish the validity and reliability of MELTEC. The following section discusses the area of analysis performed in the present study.

II. Methodology

In the present study, the prototype MELTEC was refined by fifteen professional and lay experts using the Fuzzy Delphi Method (FDM) before employing the Rasch Measurement Model. These experts validated a total of 204 items. After the validation process, the number of items was reduced to 141. During the analysis, 63 items did not meet the requirements of the Fuzzy Delphi Method. On that account, these items were dropped. All 141 items were then refined to ensure their face validity. For this purpose, ten professional and lay experts were appointed to assess the prototype checklist's overall presentation and to determine if the respondents of the study will easily understand the items. The items were refined accordingly before the Pilot study was conducted. The prototype Malaysian English Language Textbook Evaluation Checklist distributed during the Pilot study consisted of 141 items anchored in a 7-point Likert scale (1= Strongly Disagree, 2= Disagree, 3= Somewhat disagree, 4= Neither Agree nor Disagree, 5= Somewhat agree, 6=Agree, 7= Strongly Agree). The 7-point Likert scale was chosen due to the accuracy and precision of data obtained (Jamil et al., 2019). The prototype checklist was distributed to 110 ESL secondary school teachers in Kuala Lumpur, Malaysia. These ESL secondary school teachers analysed the Pulse 2 ELT textbook by Macmillan Press using the prototype checklist. They were given one week to complete the checklists, after which the checklists were collected for analysis using the Rasch Measurement Model.

III. Findings

To establish the validity and reliability of MELTEC using the Rasch Measurement Model, the analyses of item fit, item polarity, reliability, separation index and unidimensionality were performed using WINSTEPS 3.71 software. Table 1 presents the rule of thumb set for analysing the items in this study. The items that do not meet the rule of thumb set will be dropped. The remaining items will make up MELTEC.

TABLE 1
RULE OF THUMB SET FOR ITEM ANALYSIS USING RASCH MEASUREMENT MODEL (RMM)

Statistical Information	Rule of thumb
Item Fit MNSQ (Infit and Outfit)	0.6-1.4 (Bond & Fox, 2015).
Item Polarity (PT- Measure Correlation)	0.3-0.8 (Bond & Fox, 2015).
Unidimensionality Principal Component Analysis (PCA) Noise	> 20% (Reckase, 1979) < 10% (Fisher, 2007)
Reliability Item Person	> 0.8 (Fisher, 2007)
Separation Index Item Person	> 2.0 (Linacre, 2002)

A. Item Fit and Polarity

In the present study, the Infit and Outfit Mean Square Analysis (MNSQ) were analysed to measure the item fit of items in the prototype checklist. The infit and outfit MNSQ values should be between 0.6 and 1.4 (Bond & Fox, 2015). Also, the Point Measure Correlation (PTMEA Corr.) values were examined to detect the extent to which the constructs in the checklist achieve their intended goal. Therefore, to determine if the items in the checklist measure the constructs, the PTMea Correlation values must be positive and between the value of 0.3 and 0.8 (Bond & Fox, 2015). Table 2 displays the Infit, Outfit Mean Square Analysis (MNSQ) and Point Measure Correlation (PTMEA Corr.) values of all the 141 items in the prototype Malaysian English Language Textbook Evaluation Checklist. These items were first analysed for item fit. After the analysis for item fit was performed, the analysis for item polarity was performed. The items that did not fit the rule of thumb for item fit and item polarity were dropped. Based on the analyses performed, 63 items were not in the specified range of the analyses performed. Hence, these items were dropped.

TABLE 2
ITEM STATISTICS: ITEM FIT AND POLARITY

Item	Infit (MNSQ)	Outfit (MNSQ)	PT- Measure Corrrelation	Result
1	.77	.81	.16	Dropped
2	.76	.78	.24	Dropped
3	1.22	1.23	.22	Dropped
4	1.32	1.42	.51	Dropped
5	1.36	1.45	.51	Dropped
6	1.35	1.48	.49	Dropped
7	.55	.55	.62	Dropped
8	1.59	1.68	.22	Dropped
9	1.65	1.76	.29	Dropped
10	1.79	1.93	.27	Dropped
11	.72	.76	.51	Retain
12	1.31	1.57	.28	Dropped
13	.82	.73	.48	Retain
14	1.53	1.68	.37	Dropped
15	1.35	1.51	.37	Dropped
16	1.06	1.33	-.02	Dropped
17	1.35	1.53	.21	Dropped
18	.63	.64	.52	Retain
19	.67	.67	.49	Retain
20	1.12	1.27	.48	Retain
21	1.42	1.73	.33	Dropped
22	1.03	1.10	.40	Retain
23	1.43	1.78	.36	Dropped
24	1.42	1.73	.33	Dropped
25	.80	.85	.55	Retain
26	.99	1.10	.45	Retain
27	.84	.79	.55	Retain
28	1.36	1.66	.45	Dropped
29	1.26	1.55	.44	Dropped
30	1.25	1.59	.39	Dropped
Item	**Infit (MNSQ)**	**Outfit (MNSQ)**	**PT- Measure Corrrelation**	**Result**
31	1.07	1.36	.46	Retain
32	.44	.44	.52	Dropped
33	.75	.66	.67	Retain
34	.64	.57	.58	Dropped
35	.87	.86	.48	Retain

36	1.25	1.59	.39	Dropped
37	1.14	1.34	.38	Retain
38	1.36	1.45	.51	Dropped
39	.90	.98	.48	Retain
40	.95	1.05	.56	Retain
41	.81	.85	.65	Retain
42	.87	.87	.54	Retain
43	.87	.91	.53	Retain
44	1.14	1.51	.39	Dropped
45	.83	.76	.63	Retain
46	.70	.68	.57	Retain
47	.68	.61	.59	Retain
48	.71	.60	.55	Retain
49	.64	.57	.58	Dropped
50	.93	.89	.55	Retain
51	1.61	1.75	.44	Dropped
52	1.52	1.64	.41	Dropped
53	.86	.80	.47	Retain
54	.63	.64	.52	Retain
55	.95	1.09	.46	Retain
56	.73	.67	.57	Retain
57	.55	.53	.61	Dropped
58	.91	1.09	.48	Retain
59	.94	1.16	.50	Retain
60	.79	.77	.59	Retain
61	.67	.71	.65	Retain
62	.85	.87	.54	Retain
63	.74	.81	.57	Retain
64	1.41	1.53	.53	Dropped
65	1.36	1.45	.51	Dropped
66	1.06	1.08	.43	Retain
67	1.21	1.43	.33	Dropped
68	1.27	1.45	.37	Dropped
69	.88	.98	.29	Dropped
70	1.27	1.41	.42	Dropped
71	.85	.92	.54	Retain
072	.88	.87	.47	Retain
73	.76	.75	.53	Retain
74	1.03	1.11	.36	Retain
75	.72	.78	.56	Retain
76	1.79	1.93	.27	Dropped
77	.65	.62	.48	Retain
78	.66	.66	.39	Retain
79	.67	.68	.54	Retain
80	.65	.62	.49	Retain
81	.89	.95	.30	Retain
82	.65	.64	.41	Retain
83	1.75	1.92	.09	Dropped
84	.79	.86	.36	Retain
85	1.35	1.53	.29	Dropped
86	1.36	1.54	3.7	Dropped
87	.67	.67	.43	Retain
88	.82	.85	.47	Retain
89	1.36	1.54	.26	Dropped
90	1.27	1.41	.42	Dropped
91	.71	.83	.32	Retain
92	.64	.76	.41	Retain
93	.66	.72	.41	Retain
94	.68	.75	.43	Retain
95	.58	.61	.43	Dropped
96	.55	.53	.61	Dropped
97	.80	.77	.44	Retain
98	.75	.73	.39	Retain
99	.80	.76	.50	Retain
100	.78	.78	.55	Retain
Item	**Infit (MNSQ)**	**Outfit (MNSQ)**	**PT- Measure Corrrelation**	**Result**
101	1.60	1.71	.23	Dropped
102	1.43	1.56	.23	Dropped
103	1.26	1.32	.27	Dropped
104	1.43	1.56	.23	Dropped
105	.94	1.16	.50	Retain

106	.86	.87	.56	Retain
107	.88	.93	.51	Retain
108	.88	.93	.24	Dropped
109	.60	.55	.60	Dropped
110	.55	.53	.61	Dropped
111	.95	1.01	.38	Retain
112	.55	.52	.45	Dropped
113	.47	.47	.50	Dropped
114	.51	.51	.43	Dropped
115	.59	.54	.46	Dropped
116	.55	.57	.53	Dropped
117	.51	.51	.43	Dropped
118	.90	1.03	.35	Retain
119	.94	.91	.35	Retain
120	.94	.94	.48	Retain
121	.83	.80	.40	Retain
122	.86	.84	.36	Retain
123	.83	.82	.43	Retain
124	.78	.77	.42	Retain
125	1.09	1.12	.30	Retain
126	1.12	1.27	.48	Retain
127	1.35	1.53	.21	Dropped
128	.85	.97	.37	Retain
129	.44	.44	.52	Dropped
130	.71	.82	.42	Retain
131	.72	.78	.47	Retain
132	.68	.69	.43	Retain
133	1.25	1.59	.39	Dropped
134	1.36	1.45	.51	Dropped
135	.67	.73	.33	Retain
136	.81	.84	.45	Retain
137	.81	.90	.41	Retain
138	.62	.64	.48	Retain
139	1.38	1.49	.52	Dropped
140	1.36	1.45	.51	Dropped
141	.80	.80	.47	Retain

B. Reliability and Separation Index

Table 3 shows the person reliability and separation index values. The person reliability value obtained was 0.96. The value of person reliability of more than .94 indicates excellent reliability (Fisher, 2007). This indicates that MELTEC would remain consistent if given to a new set of samples with the same characteristics as the samples in this study (Kamis et al., 2015). Meanwhile, the value of person separation index was 4.86 which is higher than 2, therefore, it is acceptable (Linacre, 2002; Fox & Jones, 1998). This value indicates the levels of capabilities identified in the sample group. On the other hand, based on Table 4, the item reliability obtained was 0.92. Fisher (2007) considers it to be very good. The value of the item separation index was 3.31, which exceeded the cut-off point of 2 (Linacre, 2002), indicating good separation of item difficulty level. On the whole, the values obtained prove that MELTEC is effective with a high level of reliability in terms of person and item.

TABLE 3
PERSON RELIABILITY AND SEPARATION INDEX

Total Score		**Count**	**Measure**	**Model Error**	**Infit**		**Outfit**	
					MNSQ	**ZSTD**	**MNSQ**	**ZSTD**
MEAN	517.7	140.6	.71	.12	1.08	-.2	1.04	-.5
S.D	49.0	.9	.72	.02	.80	3.7	.81	3.8
MAX.	633.0	141.0	3.12	.16	6.14	9.9	6.01	9.9
MIN.	313.0	134.0	-1.30	.09	.14	-8.2	.12	-8.6
REAL RMSE .15		TRUE SD .71		SEPARATION 4.86			Person RELIABILITY .96	
MODEL RMSE .12		TRUE SD .71		SEPARATION 5.74			Person RELIABILITY .97	
S. E. OF Person MEAN = .07								

TABLE 4
ITEM RELIABILITY AND SEPARATION INDEX

Total Score		Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	410.9	108.9	-.12	.14	.82	-1.1	.86	-.8
S.D	29.2	.6	.50	.02	.13	.8	.18	1.1
MAX.	450.0	110.0	1.55	.18	1.14	.9	1.36	2.0
MIN.	285.0	107.0	-1.35	.10	.62	-2.6	.60	-2.4
REAL RMSE .14		TRUE SD .47		SEPARATION 3.31		Item RELIABILITY .92		
MODEL RMSE .14		TRUE SD .48		SEPARATION 3.32		Item RELIABILITY .92		
S. E. OF Item MEAN = .06								

C. Unidimensionality

In order to determine the measurement alignment of constructs, the unidimensionality features are analysed (Maat et al., 2018). The raw variance explained by measures and unexplained variance in the first construct were considered for this study. The analyses revealed that the value of raw variance explained by measures was 29.4%. It exceeded the minimum requirement of 20% (Reckase, 1979). Also, the unexplained variance in the 1st construct was 8.8%. The value is proven to be less than 15% which indicates a good value (Fisher, 2007). Hence, the Malaysian English Language Textbook Evaluation Checklist items are proven to be accurate and fit the intended purpose of evaluating ELT textbooks.

TABLE 5
RESIDUAL PRINCIPAL COMPONENT ANALYSIS (PCA)

Raw Variance Explained By Measures	Unexplained Variance In 1st Contrast
29.4%	8.8%

IV. DISCUSSION

After performing the data analyses in the present study following the benchmark and conditions of Rasch Measurement Model, 63 items in the prototype checklist were excluded. These items did not perform their function to measure the variables in this study. After excluding these items, the finalised version of the Malaysian English Language Textbook Evaluation Checklist consists of 78 items. The use of Rasch Measurement Model in the present study provided empirical evidence to prove the quality of items in the checklist. The summary of items dropped and retained is shown in Table 6.

TABLE 6
SUMMARY OF ITEMS DROPPED AND RETAINED

Phase	No. of items	Items Dropped
1 (Design and Development)	204	0
2 (Validation)	141	63
TOTAL	78	-

With the emergence of the Rasch Measurement Model, significant contributions to various fields of research have been recorded (Aryadoust, Tan & Ng, 2019). However, the application of the Rasch Measurement Model in the field of ELT textbook evaluation, particularly in the development and validation of ELT textbook evaluation checklists is non-existent. Previous textbook evaluation checklists have utilised the Critical Test Theory (CTT) to determine the validity and reliability of the instruments developed (Nguyen, 2015; Isik, 2018; Simsek & Dündar, 2018; Nimehchisalem & Mukundan, 2015; AbdelWahab, 2013; Karamifar, Barati, & Youhanaee, 2014; Lopez Medina, 2016; Zokaeieh et al., 2019; Sahin, 2020). None up to date studies have considered employing Rasch Measurement Model, an Item Response Theory (IRT), to establish the psychometric properties of the items and the validity and reliability of the instruments developed. As aforementioned, textbook evaluation checklists in the past lack validity and reliability (Nimehchisalem & Mukundan, 2015). Hence, by employing the Rasch Measurement Model in the present study, the validity and reliability of the Malaysian English Language Textbook Evaluation Checklist (MELTEC) have been established.

V. CONCLUSION

The analyses performed employing the Rasch Measurement Model to establish the validity and reliability of the Malaysian English Language Textbook Evaluation Checklist (MELTEC) have bridged the gap between research and practice. The examination of validity and reliability in this study provides an alternative for future researchers in the field of ELT textbook evaluation to consider establishing the validity and reliability of instruments developed using the Item Response Theory (IRT) instead of the Critical Test Theory (CTT) model of measurement. The aspects of validity and reliability are undoubtedly essential and should not be compromised, especially when developing a new research instrument. In essence, the assumptions fulfilled in this study using the Rasch Measurement Model were; Item Fit,

Polarity, Reliability, Separation Index, and Unidimensionality. Through this paper, researchers especially in the language area can develop a new understanding of the integration of Rasch Measurement Model which is relatively under-studied especially in the field of ELT textbook evaluation. Furthermore, there are a total of eight diagnostic data analyses that can be performed using the Rasch Measurement Model, which includes (i) unidimensional; (ii) compatibility (fit) item; (iii) polarities item; (iv) reliability and separation item respondents; (v) appropriateness of the measurement scale based on the use of categories; (vi) value of standardised residual correlation in determining leaning item; (vii) differential items functioning (DIF) based on gender; and (viii) the distribution of item difficulty levels and abilities of respondents. Hence, future researchers may consider exploring other assumptions of the Rasch Measurement Model not explored in this study that would suit the objectives and purpose of their study.

ACKNOWLEDGMENTS

The researchers would like to thank the National University of Malaysia (UKM) for its support for this research via its Grant Tabung Agihan Penyelidikan (TAP-K017393).

REFERENCES

[1] AbdelWahab, M. M. (2013). Developing an English Language Textbook Evaluative Checklist. *Journal of Research & Method in Education*, *1*(3), 55-70.
[2] Aryadoust, V., Tan, H. A. H., & Ng, L. Y. (2019). A Scientometric Review of Rasch Measurement: The Rise and Progress of a Specialty. *Frontiers in Psychology*, *10*(2197), 1–16.
[3] Bond, T. G., & Fox, C. M. B. (2015). *Applying the Rasch: Fundamental measurement in the Human Sciences* (Third Edition). New York: Routledge.
[4] Cunningsworth, A. (1995). *Choosing Your Coursebook (Handbooks for the English Classroom)*. Oxford: Macmillan Heinemann.
[5] Demir, Y., & Ertas, A. (2014). A suggested eclectic checklist for ELT coursebook evaluation. *The Reading Matrix*, *14*(2), 243–252.
[6] Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095.
[7] Fox, C. M., & Jones, J. A. (1998). Uses of Rasch Modeling in Counseling Psychology Research. *Journal of Counseling Psychology*, *45*(1), 30–45.
[8] Halim, S., & Halim, T. (2016). Adapting Materials: Revisiting the Needs of Learners. *International Journal of Humanities and Cultural Studies*, *2*(4), 633–642.
[9] Hassan, A. B. (2012). *Instrumen Penilaian Pembimbing Dalam Pelaksanaan Pembelajaran Berasaskan Kerja Pelajar Di Industri*. [Unpublished Ph.D. dissertation], Universiti Teknologi Malaysia.
[10] Işik, A. (2018). ELT Materials Evaluation : A System and Criteria. *Theory and Practice in Language Studies*, *8*(7), 797–812.
[11] Jamil, M. R. B. M., Siraj, S. B., Hussin, Z. B., Noh, N. R. B. M., & Sapar, A. A. B. (2019). *Pengenalan Asas Kaedah Fuzzy Delphi: Dalam Penyelidikan Reka Bentuk Pembangunan* (Third Edit). Minda Intelek Agency.
[12] Kamis, A., Bakar, A. R., Hamzah, R., Soaib A., & Zaliza, H. (2015). Using Rasch model and confirmatory factor analysis to assess instrument for clothing fashion design competency. *International Journal of Social Science and Humanity*, *5*(5), 418–421.
[13] Karamifar, M., Barati, H., & Youhanaee, M. (2014). Local Evaluation Criteria for Global Textbooks: A Case Study from Iran. *Theory and Practice in Language Studies*, *4(*5), 923–930.
[14] Linacre, J. (2002). Understanding Rasch measurement: Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, *3*, 85–106.
[15] Lopez Medina, B. (2016). Developing a CLIL textbook evaluation checklist. *Latin American Journal of Content & Language Integrated Learning*, *9*(1), 159–173.
[16] Maat, S. M., Effendi, M. E., Affandi, H. M., & Mohamed, S. (2018). Psychometric evidence of statistical self-efficacy instrument based on postgraduate. *International Journal of Advanced and Applied Sciences*, *5*(11), 95–99.
[17] McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and Methods in ELT: A Teacher's Guide* (Third Edit). West Sussex: Wiley-Blackwell.
[18] McGrath, I. (2002). *Materials Evaluation and Design for Language Teaching*. Edinburgh: Edinburgh University Press.
[19] Mukundan, J., & Ahour, T. (2010). A review of Textbook Evaluation Checklist across Four Decades (1970-2008). In Tomlinson, B. & Mahasura, H. (Eds.). *Research for Materials Development in Language Learning: Evidence for best practice* (pp. 336–352). Continuum International Publishing.
[20] Nehal, R. (2016). Materials Adaptation in English Language Teaching : Implications for Teacher Development. In Rahman, M. M. (Ed.) (2016). *Integrating Technology and Culture: Strategies and Innovations in ELT* (pp. 239–262), Anubhav Publishing House: India.
[21] Nguyen, C. (2015). *An Evaluation of the Textbook English 6 : A case study from secondary schools in the Mekong Delta Provinces of Vietnam*. [Unpublished Ph.D. dissertation], University of Sheffield.
[22] Nimehchisalem, V., & Mukundan, J. (2013). Usefulness of the English Language Teaching Textbook Evaluation Checklist. *Pertanika Journal of Social Sciences and Humanities*, *21*(2), 797–816.
[23] Nimehchisalem, V., & Mukundan, J. (2015). Refinement of the English Language Teaching Textbook Evaluation Checklist. *Pertanika Journal of Social Sciences and Humanities*, *23*(4), 761–780.
[24] Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational and Behavioral Statistics*, *4*(3), 207–230.
[25] Sahin, S. (2020). Developing a Checklist for English Language Teaching Course book Analysis. *International Journal of*

Education and Research, 8(1), 107–120.
[26] Simsek, M., & Dündar, E. (2018). Particularised Checklists in Materials Evaluation: Developing Contextually Relevant Criteria for Turkish EFL Classes. *Journal of Language and Linguistic Studies, 14*(3), 154–189.
[27] Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching, 45*(2), 143–179.
[28] Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. MESA Press: Chicago, IL
[29] Zokaeieh, S., Karimi, S. N., Nouri, F., & Hakimzadeh, A. (2019). Developing a Critical Checklist for Textbook Evaluation. *The Journal of English Language Pedagogy and Practice, 12*(24), 202–225.

Faustina Roberts is currently pursuing her Ph.D. (TESL) at the Faculty of Education, National University of Malaysia. Her research interests include ELT materials development, textbook evaluation, and instrument development. Email: tinaroberts716@gmail.com

Azlina Abdul Aziz is a TESL lecturer at the Faculty of Education, National University of Malaysia. She has an Ed.D in Teaching of English from Teachers College, Columbia University, U.S.A. Her research interests include Teaching and Learning of Literature and Teacher Education in TESL. She is also interested in how literary texts and personal narratives may help students examine the social, cultural, and political issues in a particular context. Email: azlina1@ukm.edu.my

Mohd Effendi @ Ewan Mohd Matore is a senior lecturer at the Faculty of Education, National University of Malaysia. He obtained his Ph.D. (Psychometric and Evaluation) from Universiti Sains Malaysia, USM, Malaysia. His research interests include psychometric and evaluation, educational measurement, and instrument development. Email: effendi@ukm.edu.my