

JIRSEA Issue: Vol. 22 No. 3

Sept/Oct 2024

ISSN 1675-6061

<http://www.seairweb.info/journal/index.aspx>

Submission Timeline

First submission: 19 May 2024

Revised submission: 24 July 2024

Final Submission: 25 October 2024

Acceptance: 29 October 2024

Evaluating an ESL Textbooks Evaluation Questionnaire using Rasch Analysis

Geethaletchmi Ponniah, Samah Ali Mohsen Mofreh, and Sultan Salem

CITE ARTICLE: Ponniah, G., Samah Ali Mohsen Mofreh, S.A.M., and Salem, S. (2024). *Journal of Institutional Research South East Asia*, 22(3), 192-210

DECLARATION: *Articles published in JIRSEA adhere to the declaration clauses that: (1) the article is entirely the researcher(s) original work and responsibility, (2) principles of honesty and integrity throughout the research process, and unless otherwise indicated and properly cited, this declaration encompasses all aspects of the article, including but not limited to text, figures, tables, data, and any accompanying material are observed, (3) there is no conflict of interest with JIRSEA in any form, (4) holding to all required JIRSEA policies of publication ethics and practices conforming to the COPE Principles of Transparency and Best Practice in Scholarly Publishing, (5) copyrights assignment to JIRSEA, whereby it is published online in an open access requirement of Creative Commons 4.0, and (6) is not funded by any party unless otherwise disclosed.*



Publisher: SEAIR Secretariat

C/O Suan Dusit Rajaphat University
295 Nakhon Ratchasima Rd, Dusit, Dusit
District, Bangkok 10300, THAILAND
email:seair.info@gmail.com
<http://www.seairweb.info/>

CREATIVE COMMONS ATTRIBUTION 4.0



INTERNATIONAL LICENSE

EVALUATING AN ESL TEXTBOOKS EVALUATION QUESTIONNAIRE USING RASCH ANALYSIS

Geethaletchmi Ponniah*¹, Samah Ali Mohsen Mofreh ², Sultan Salem ³

¹ *School of Educational Studies, Universiti Sains Malaysia, *Corresponding author:*
gp14_edu086@student.usm.my

² *School of Educational Studies, Universiti Sains Malaysia*

³ *Department of Economics, Birmingham Business School,
College of Social Sciences, University of Birmingham*

ABSTRACT

A valid, reliable, and practical instrument is needed to evaluate ESL textbooks in Malaysian community colleges. Thus, a questionnaire guided by the Litz Theory (2005) was designed to evaluate ESL textbooks in Malaysian community colleges. This paper assessed the validation of the questionnaire using the Rasch Model analysis of construct validity and reliability. The questionnaire investigates community college students' perceptions of the ESL textbooks they use in their classrooms. The sample of this study comprised 123 community college students in Malaysia. The Rasch Model produced construct validity and reliability measurements using the dimensionality, item fit, and item polarity parameters and the person and item separation analysis. The results of the Rasch Model Analysis reveal that the construct validity in the study shows the uniformity of the instruments in the good category. This indicates that the items used in this study are related to the material's content. Hence, this questionnaire is valid and reliable to evaluate ESL textbooks in Malaysian community colleges.

Keywords: Evaluation, textbook, textbook evaluation, Rasch model

Introduction

Textbooks are packages with different but interrelated parts. They are the main sources that could convey knowledge and information to the learners in an easy and organized way (Ahour & Ahmadi, 2012; Orfan & Akramy, 2021). According to Prabhu (1987), textbooks are fully specified and preconstructed materials that provide a certain amount of uniformity in what occurs in many different classes with different teachers and students, which serves the interests of accountability. Despite the emergence of recent technologies in the field of education, copies of textbooks are still the most commonly used source material for most instructional situations as it is for language teaching contexts. On the significance of course books, Hutchinson and Torres (1994, as cited in Litz, 2005) suggest that “the textbook is an almost universal element of [English language] teaching. Millions of copies are sold every year, and numerous aid projects have been set up to produce those in [various] countries. No teaching-learning situation, it seems, is complete until it has its relevant textbook” (p. 315). There are different views about the textbook. Cunningsworth (1995), for example, identifies a textbook as a source for presenting the materials and learners to practice and do the activities. Hutchinson and Torres (1994) argue that textbooks have a very important and assertive role in teaching and learning (Gholampour & Mehrabi, 2023).

Sheldon (1988) suggests that “textbooks do not only represent the visible heart of any English Language Teaching program but also offer considerable advantages” (p. 237). The most essential function of a textbook is to motivate students to learn (Liu & Diana Deris, 2023). For Dubin and Olshtain (1986), “the tangible element that gives language course face validity to many teachers and learners is the textbook” (p.167) (Liu & Deris, 2023). Ur (1996) believes that a textbook provides a clear framework. It clarifies what is coming next, and learners know where they are going. Concerning the essential aim of ELT textbooks, Byrd (2001) argues that ELT textbooks include two kinds of information, which are topic content (e.g., family, school, etc.) and linguistic content (e.g., grammar, vocabulary, and skills) and that they help learners to learn the latter from the former. Considering the importance of textbooks, evaluating a book needs a significant interest because it provides useful information for teachers but also leads to helping students in learning settings.

Developing the researcher’s instrument requires knowledge about item or question construction, scale development, format, length, validity, and reliability of the instrument and its scores (Sekaran, 2003; Creswell, 2012; Johnson & Christensen, 2012). Many instruments to evaluate ESL textbooks are available in the literature; however, the problem is that most of them have not been validated. Although some instruments have been tested for their validity and reliability, they lack stronger evidence of psychometrics in construct validity. While some are too long, which reduces their economy, others are too sophisticated, which makes it very challenging for novice lecturers to use them. Therefore, there is a need for an instrument that is valid, reliable, and practical in the parlance of psychometrics. Thus, this study aims to evaluate the validity and reliability of the newly developed questionnaire to evaluate ESL textbooks in Malaysian community colleges using the Rasch Measurement Model (RMM). In particular, the study's objective was to develop a measurable questionnaire to evaluate the students’ feedback on the suitability of ESL textbooks used at Community Colleges in Malaysia.

Literature Review

ESL Textbook in Malaysian Community Colleges

Currently, Malaysia has some critical issues with ESL. One of the major issues is that it has to find the best method of teaching English to multilingual students. When selecting textbooks, the needs of students must be considered. Students' needs encompass both personal learning and future professional needs. This issue can be solved by selecting a suitable textbook that fulfills the students' needs (Maftukah & Astuti, 2021). Unfortunately, this issue remains unsolved even though the quality of ESL textbooks has improved dramatically in recent years. This is because the lecturers are influenced by publishers' representatives who may provide informed assistance to sell their products (Maftukah & Astuti, 2021).

The latest problem faced by the lecturers and students of community colleges is the lack of core textbooks. This is because according to the English syllabus of Community Colleges, the core textbook recommended by MQA is Collins Easy Learning English Conversation Book 1 (2015): United Kingdom: Harper Collins Publishers. However, it is not available since it is out of publication. So, the lecturers of community colleges are using other reference books to conduct their English lessons. One of the reference books used is Practice Makes Perfect: English Conversation. It is a workbook written by Yates. Even though the particular workbook provided practical exercises for the students to practice, it fails to provide sufficient theory since it is a workbook, not a textbook that usually contains detailed theory concepts. This also leads the lecturers to spend more time preparing their lessons in terms of theory since the theory part is lacking in the workbook.

To improve the quality of teaching and learning in higher education institutions (HEI), the National Higher Education Strategic Plan (NHESP) states that HEIs should conduct curriculum reviews every two or three years, taking into account the perspectives of academics, industry experts, government officials, and members of non-governmental organizations (Weiss, 2014). Moreover, it was emphasized that curricular transformation should aim for English language proficiency (Ramamurthy et al., 2020). Thus, lecturers are encouraged to review and revise existing teaching modules used as textbooks in classrooms following industry requirements (Amin, 2016; Chinedu & Mohamed, 2017; Rasul et al., 2015 & Wijayanto, 2017; Ramamurthy et al., 2020). This will provide new opportunities for them to effectively deliver the teaching modules used as textbooks using modern community colleges' methods and concepts (Grosch, 2017; Ramamurthy et al., 2020). Therefore, lecturers should also create teaching modules used as textbooks that are tailored to the industry that the institution wants its graduates to work in after they complete their studies. This will improve their employability because graduates capable of communicating technical information within the industry are regarded as all-rounders. Furthermore, lecturers are encouraged to collaborate with industry experts to design educational programs based on industry input (Grosch, 2017; Ramamurthy et al., 2020). These statements clearly show the importance of choosing a suitable textbook for classroom use. Therefore, to obtain the most suitable textbook, the lecturers must know the strategies and criteria for selecting suitable textbooks.

Many studies have also shown that textbook selection is an important aspect of learning activity (e.g. Lemmer, 2008; Mijayanti, 2015; Işık & Kurum, 2002; Sikorova, 2004; Watt, 2009; Ho & Hsu, 2011; Yuen & Ting, 2012; Amerian and Khaiyar, 2014; Marczak, 2013; Chang, 2002; Huang, 2011; Kim, 2002; Garinger, 2002; Kiai & Maroko, 2013; Fredriksson &

Olsson, 2006; Mahmood, 2011; LaBelle, 2010; Deuri, 2012; Maftukah & Astuti, 2021). For the criteria of an appropriate English textbook, researchers discovered a plethora of criteria used by teachers in selecting their textbook (e.g., Rahimpour & Hashemi, 2011; Makgato & Ramaligela, 2013; Rodriguez, 2015; Reid, 2017; Mukundan, 2011; Akbar, 2016; Shabani, 2017; Maftukah & Astuti, 2021).

However, lecturers at community colleges continue to choose inappropriate textbooks because they are unaware of the proper strategies and criteria for selecting appropriate ones to obtain the best ones. Besides, very limited studies concentrate on strategies such as how textbooks are chosen, the steps involved in selecting an appropriate English textbook, and the criteria for selecting an appropriate textbook. The gaps revealed that the research covers a broad range of topics. Most previous studies conducted their research without focusing on a single institution (Maftukah & Astuti, 2021). To limit the scope of the study, it was critical to conduct the study with a focus on a specific institution. Therefore, this study will concentrate on community colleges in Malaysia.

Students' Views and Textbook Evaluation

Earlier research suggested that authors and publishers of textbooks ought to consider students' perspectives during the evaluation process (Anderson, 1989; McDonough et al., 2017). In this context, McDonough et al. (2017) critique studies where researchers gather lecturers' views on students' needs, as these may not accurately represent the true needs of the students. Kumaravadivelu (2006) discusses the concept of "dramatic mismatches" (p. 106) that arise between the perspectives of lecturers and students. Additionally, Preedy (2001) outlines four reasons for including students in the textbook evaluation process. Firstly, students have distinct perspectives on textbooks compared to lecturers and other parties involved. As a result, students can offer insights about textbooks drawn from their own experiences and provide valuable feedback. Secondly, including students in the evaluation process boosts their self-efficacy. This increased sense of efficacy fosters greater motivation towards using these materials and can enhance their dedication to learning (Preedy, 2001). Thirdly, there is considerable theoretical evidence suggesting that discrepancies between students' and lecturers' views on learning and educational resources can create significant barriers when establishing language learning goals and objectives in classrooms (Emelyanova & Voronina, 2014; Van, 2011; Winne & Marx, 1982). Finally, involving students in the evaluation of textbooks promotes reflection and autonomy, encouraging students to take charge of their learning. Thus, this research will concentrate on students' perspectives regarding assessing the ESL textbook's appropriateness.

Checklist as the Method of Textbook Evaluation

There are several techniques for evaluating textbooks. A criterion-based checklist is one of the most used (Abdel Wahab, 2013; McGrath, 2016; Richards, 2016). In the context of English language teaching (ELT), Mukundan et al. (2011) define a checklist as an instrument that assists practitioners in evaluating materials such as textbooks for English language acquisition. According to Brown (2001), textbook evaluation checklists include a thorough list of criteria that enable the review process to be completed methodically. These characteristics could include a textbook's physical appearance, tasks, exercises, and activities, language skill coverage (reading, writing, listening, speaking, grammar, vocabulary), alignment with a syllabus and curriculum, and learner compatibility. There are numerous reasons why ELT textbook evaluators worldwide utilize evaluation checklists. For example, they allow for extensive and in-depth examination, especially when applying qualitative

measurements (Cunningsworth, 1995; Skierso, 1991; Mukundan et al., 2011; Demir & Ertas, 2014). A checklist can also be simply duplicated (Ellis, 1997) and tailored to the demands of future users (Mukundan & Ahour, 2010). Finally, it is thought to be cost-effective, as it allows for recording a large amount of information in a short period. Thus, a questionnaire guided by the Litz (2005) checklist model was designed to evaluate ESL textbooks in Malaysian community colleges.

Instrument and Materials

Litz (2005) Checklist Model

Litz's (2005) checklist model created a set of questionnaires for evaluating textbooks that are designed for both instructors and students to fill out. The questions cover several aspects: a) practical considerations, b) layout and design, c) activities, d) skills, e) language type, f) subject and content, and g) conclusion/overall consensus. Each of these categories includes specific evaluative items for the student and teacher evaluation forms: a) 2 and 5, b) 2 and 8, c) 5 and 7, d) 3 and 5, e) 6 and 6, f) 5 and 5, and g) 2 and 4. The rating scale employs a 10-point system ranging from 1 (Highly Disagree) to 10 (Highly Agree). Litz's checklist is structured to highlight several primary categories along with detailed items. The items use straightforward language and are generally easy to understand, making the checklist thorough and well-balanced regarding the various topics it addresses. A crucial aspect is that the checklist was piloted simultaneously, meaning the author provided descriptions that can assist users in understanding which features to consider.

Litz's (2005) checklist model was employed in this research as it serves as a reliable metric, being a standardized checklist recognized globally for evaluating books and is regarded as the most commonly adapted textbook evaluation tool, referenced by six studies (Ahour et al., 2014; Khodabakshi, 2014; Ghezlou et al., 2016; Nourmohammad-Nouri et al., 2015; Monazzah et al., 2016; Ahmad et al., 2019). The newly developed textbook evaluation questionnaire emphasizes eight criteria: student demographics, practical considerations, layout and design, activities, skills, language type, subject matter, and overall consensus. The elements of Litz's (2005) model checklists were structured on a 10-point scale, ranging from 1 (Highly Disagree) to 10 (Highly Agree). For the current research, these 10 scales were simplified to a 5-point Likert Scale for statistical definition and coding (e.g., Strongly Disagree = 1, Disagree = 2, Neutral = 3, Agree = 4, Strongly Agree = 5). This study utilized an evaluation questionnaire (ESL Textbook Evaluation Questionnaire) based on Litz's (2005) checklist model to gather information regarding students' views on the ESL textbook. Consequently, the questionnaire was exclusively designed for student input. The student questionnaire included 41 items.

Questionnaire To Evaluate ESL Textbooks

To gather the required data, a questionnaire was developed to evaluate ESL textbooks following the Litz (2005) checklist model. Two English experts validated the developed questionnaire to evaluate ESL textbook content. The developed questionnaire to evaluate ESL textbooks consisted of sections 1 and 2. Section 1 was used to gather information on the student demographics, which include gender, age, level of education, and cultural background (i.e., Malaysian or others). Section 2 carries 37-point questions categorized into seven sections, namely, practical considerations, layout and design, activities, skills, language types, subject and content, and overall consensus. The developed scale to evaluate ESL textbooks is

in Likert scale format, comprised of “Strongly disagree (1)”, “Disagree (2)”, “Neutral (3)”, “Agree (4)” and “Strongly agree (5)”.

Table 1 summarizes the dimensions of the questionnaire developed based on the Litz (2005) Checklist Model.

Table 1: Specification

Section	Dimensions	Operational Definitions	# of Items
1	Students’ Demographics	This section covers gender, age, educational background, and race.	4
2 a	Practical considerations	This section focuses on pricing and accessibility.	2
2b	Layout and design	This section includes the layout of the textbook and the content page.	2
2c	Activities	This section pertains to all the activities and tasks featured in the book.	13
2d	Skills	This part evaluates four skills such as listening, speaking, reading, and writing.	3
2e	Language types	This section evaluates language types, such as realistic and authentic.	7
2f	Subject and content	This section addresses all the subjects mentioned in the book.	8
2g	Overall consensus.	General perspective on students’ viewpoints.	2
Total Items		41	41

Item Response Theory

Item response theory (IRT) originally became popular in the 1970s, when it was used to create standardized exams like the Scholastic Aptitude exams (SATs) (Lord & Novick, 1969). IRT eventually became the most significant psychometric approach for validating scales since it addresses many of the measurement difficulties that must be addressed when developing a test or scale (Lord, 1980).

The Rasch measurement model

The Rasch Measurement Model (RMM) was used to determine the validity and reliability of the instrument items. According to Planinic et al. (2019), using the Rasch model for data analysis is a concept that has been introduced previously. The Rasch Measurement Model is an effective technique for guaranteeing the instrument's validity and reliability by providing precise data (Bond & Fox, 2015). RMM evaluated each respondent's ability to complete the instrument and measured the difficulty of each item (Green & Frantom, 2002). Furthermore, RMM may detect hidden features, including human thoughts and emotions (Planinic et al., 2019). RMM, which is based on item response theory, is a statistical model that can assess both the difficulty of the item and the skill of the person being tested (Testa et al., 2019). As a result, the RMM was able to determine both the validity and the reliability of the items and the respondents. Rasch analysis can also assess concept validity regarding item polarity, item fit, person fit, and unidimensionality. Rasch analysis takes longer than standard analysis, providing a more detailed insight into an instrument's strengths and flaws (Boone, 2016).

Methods

The research design is quantitative research. In this study, the data was collected using the newly developed ESL Textbook Evaluation Questionnaire distributed to students currently taking English courses in Community Colleges. The sample of this study was 123 students who are currently taking English courses at Community Colleges. The study participants were both male and female. The students' age ranges were between 18 and 23. The Rasch Model was used to determine the psychometric properties of the newly developed ESL Textbook Evaluation Questionnaire to test the validity and reliability of the newly developed ESL Textbook Evaluation Questionnaire.

Methodology

Rasch Model Analysis

Item fit and dimensionality analyses were conducted to test the construct validity using the Rasch Model. The reliability of instruments was conducted for this study using the Rasch Model software. The copies of the developed scale in the form of a questionnaire to evaluate ESL textbooks were distributed among the study participants. The questionnaire to evaluate the ESL textbook of this study was given to English learners of Community colleges. Before administering the questionnaire, students were given a complete explanation of the aim of the study, the developed questionnaire, and how they should be answered. The respondents were assured that the results would be used for this study and that their views would be kept completely confidential. They were given a contact number and e-mail address, and they were asked to contact the researcher if they had any questions regarding the scale. This phase of data collection was done for one month. Data were collected and prepared for statistical analysis.

Findings

The Rasch Measurement Model, which is a one-parameter Item Response Theory (IRT), is a contemporary measurement approach frequently employed in the social sciences. The limitations of analytical methods associated with Critical Test Theory (CTT), such as Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) in assessing validity and reliability, justify the selection of the Rasch Measurement Model for the current study. The Rasch Measurement Model has garnered significant interest from researchers globally, particularly in the realm of developing and constructing new assessment tools. Researchers often favor Georg Rasch's model due to its benefits, which include linearity, independence, objectivity, comprehensiveness, and the ability to draw inferences easily (Wright & Stone, 1979). This model posits that an individual's response to an item is determined solely by their ability and the item's difficulty (Bond & Fox, 2015). According to the Rasch model, each item is formulated based only on its difficulty parameter. Given a difficulty logit of 0.00, an individual has a 50% chance of answering the items correctly. Therefore, it can be understood that increasing the difficulty of an item will impact the likelihood of success, leading to a decrease in chances. The sequence of assessment within the model may differ based on the specific requirements of a study. Under the Rasch Measurement Model, eight types of diagnostic data analyses are integral to the instrument development process, which consist of (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 standardized 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 current study, the following areas of analysis are performed using the Rasch Measurement Model: (i) item polarity, (ii) item fit, (iii) unidimensionality, and (iv) reliability and separation index. These analyses are adequate for establishing the validity and reliability of the newly developed questionnaire. The following section discusses the area of analysis conducted in the current study.

Data Analysis Procedure

The study's findings were examined to establish reliability and validity in terms of construct validity. WINSTEPS software version 3.68.2 is used to assess construct validity and item reliability. This ensures that the instrument is of high quality and that the data collected by the researchers is accurate before being used in a study. To begin, PTMEA-CORR value analysis was used to identify item polarity. A positive PTMEA-CORR score indicates that the item can accurately measure what it wishes to measure, whereas a negative number suggests that it cannot. The values MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR were used to measure item fit (Bond & Fox, 2007; Boone et al., 2014; Zahid et al., 2019). The fit value of this item reflects if it can measure what it is supposed to measure (Sumintono & Widhiarso, 2015). Items not under the Item Fit Index (Table 2) must be altered or eliminated to boost the item fit value (Sumintono & Widhiarso, 2015).

Table 2: Fit Indices for Item Fit Statistics

Statistics	Fit Indices
Outfit mean square values (MNSQ)	0.50 – 1.50
Outfit z-standardized values (ZSTD)	-2.00 – 2.00
Pont Measure Correlation (PTMEA-CORR)	0.40 - 0.85

Source: Boone et al. (2014)

Researchers also assessed the instrument's unidimensionality to guarantee that it could accurately quantify instrument uniformity (Saud et al., 2018; Sumintono & Widhiarso, 2015). The principal component analysis (PCA) gives unidimensionality requirements based on 'raw variance explained by measures' (Sumintono, 2016; Sumintono & Widhiarso, 2015). The Raw variance explained by measures higher than 20% is acceptable, higher than 40% is good, and higher than 60% is excellent. Furthermore, the number for 'unexplained variance in first contrast' cannot exceed 15%. Table 3 illustrates unidimensionality based on raw variance explained by measurements.

Table 3: Unidimensionality based on Raw Variance Explained by Measures

Value	Interpretation
> 20%	Acceptable
> 40%	Good
> 60%	Excellent

Source: Sumintono and Widhiarso (2015)

As regards the reliability, the researcher refers to Sumintono and Widhiarso (2015) for, item and respondent reliability indices as well as item separation and person as shown in Table 4.

Table 4: Reliability in Rasch Analysis

Statistics	Fit Indices	Interpretation
Item and Person Reliability	<0.67	Low

	0.67-0.80	Sufficient
	0.81-0.90	Good
	0.91-0.94	Very Good
	>0.94	Excellent
Item and Person Separation	≤ 2	A high separation value indicates that the instruments are of good quality since they can identify the group of items and respondents.

Source: Sumintono & Widhiarso, 2015

Findings and Discussion

Many studies supported this study's findings by using the Rasch Model to examine construct validity (Fox and Jones, 1998; Forkmann et al., 2009; Wolfe et al., 2009; Mofreh et al., 2014). According to Mofreh et al. (2014), dimensionality, item plurality, calibration scales, and item fit analysis by the Rasch Model were used to examine the construct validity of lecturers' instructional functions instrument.

Instrument Analysis

Measuring the suitability of an ESL textbook requires a measurable instrument to be developed and tested to determine the suitability of an ESL textbook. The developed instrument was analyzed based on the research objective. This study aims to identify the constructs of validity of the ESL Textbook Evaluation Questionnaire. To achieve this objective, psychometric properties were tested for the ESL textbook evaluation questionnaire to determine if this instrument is sufficiently valid and reliable as a measurement tool. Thus, the Rasch Model analysis was used to test the validity and reliability of the ESL textbook evaluation questionnaire. In addition, the Rasch Model analysis was used to answer the research objective as described: To develop a measurable questionnaire to evaluate students' feedback on the suitability of ESL textbooks used at Community Colleges in Malaysia.

The following analysis of the Rasch Model for validity and reliability of the ESL textbook evaluation questionnaire answered the research objective related to the Rasch Model mentioned above.

Validity Analysis

Rasch Model tested the validity of the ESL Textbook Evaluation Questionnaire using item polarity, item fit, and dimensionality as psychometric properties criteria. Item polarity or point measure correlation (PTMEA CORR.) is the early detection of construct validity (Bond & Fox, 2007).

Construct Validity Results

Item Polarity

The purpose of verifying the point measure correlation value (PTMEA CORR.) is to determine the item's polarity and how well the construct's development meets its goal. According to Bond and Fox (2007), a high PTMEA-CORR value, such as 0.4–0.85, means that the item can differentiate between study subjects and that the item construction can calculate what should be measured. However, Bond and Fox (2015) state that the item calculates the desired construct if the PTMEA CORR value is positive (+). On the other side, if the result obtained is negative (-), the generated item did not calculate the measured construct. The item must then be revised or eliminated since it is unsuited to inquiries or is

difficult for the respondent to answer. The PTMEA CORR values are depicted in Figure 1. Figure 1 shows that all items have positive PTMEA-CORR values ranging from 0 to 0.75. Bond and Fox (2007) agree that a positive PTMEA-CORR value shows that the item assesses the desired construct.

Measure	Model S.E.	Infit		Outfit		PT-MEASURE		Exact OBS%	Match EXP%	ITEM
		MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
.49	.35	1.16	.6	1.09	.4	0	.26	92.7	92.7	V29_A
1.75	.19	1.37	3.2	1.43	3.2	0.04	.43	67.5	68.1	V22_A
1.26	.14	2.59	6.4	2.34	5.3	0.07	.62	67.5	66.7	V42_A
1.71	.18	1.38	3.3	1.41	3.4	0.1	.45	68.3	65.6	V20_A
-0.75	.15	1.56	3.2	1.62	3.6	0.14	.49	66.7	64.2	ACTIVITI
2.49	.20	1.14	1.5	1.39	2.7	0.15	.39	69.1	70.3	V34_A
-0.47	.12	1.85	4.9	2.08	5.4	0.17	.55	49.6	56.3	V39_A
-0.92	.25	1.22	1.2	1.30	1.3	0.18	.39	79.7	84.5	SKILLSQ2
0.79	.20	1.11	1.3	1.15	1.3	0.32	.42	67.5	70.3	V13_A
-0.57	.17	1.17	1.3	1.18	1.4	0.33	.47	63.4	64.9	V32_A
-0.92	.25	1.06	.4	1.07	.4	0.34	.39	82.9	84.5	OVERALLC
0.52	.17	1.15	1.2	1.18	1.4	0.35	.48	69.1	66.3	V40_A
-0.94	.16	1.08	.4	1.23	1.1	0.36	.45	74	72.8	SKILLSOT
-0.64	.16	1.11	.7	1.07	.5	0.4	.47	66.7	68.5	SKILLSQ3
1.64	.24	.89	.7	.89	-.5	0.44	.42	82.9	82.2	V14_A
-1.2	.16	1.05	.4	1.05	.4	0.45	.32	65	65.5	LANGUAGE
-0.89	.17	1.02	.2	1.00	.0	0.45	.48	72.4	68.9	V41_A
0.06	.20	.95	-.3	.95	-.2	0.45	.46	80.5	77.1	V46_A
0.33	.18	.92	-.5	.93	-.5	0.47	.43	75.6	71.6	V45_A
-1.08	.29	.77	-1.1	.83	-.5	0.51	.45	91.1	88.4	SUBJECTA
-0.44	.12	1.15	1.1	1.13	.9	0.53	.36	55.3	52.4	LAYOUTAN
-0.88	.17	.91	-.6	.91	-.6	0.54	.56	73.2	67.4	V44_A
0.35	.18	.90	-.7	.90	-.7	0.54	.47	68.3	69.4	V21_A
0.23	.18	.89	-.8	.89	-.8	0.54	.46	74.8	71.1	V43_A
0.31	.21	.84	-1.0	.80	-1.2	0.55	.46	78.9	77.9	V15_A
-0.94	.17	.87	-1.0	.86	-1.0	0.58	.41	69.9	66.2	V35_A
-0.26	.14	.91	-.7	.89	-.8	0.59	.48	56.9	56.4	PRACTICA
0.29	.18	.82	-1.4	.81	-1.4	0.6	.54	74.8	70.9	V37_A
-0.92	.16	.82	-1.0	.81	-1.1	0.61	.46	75.6	66.5	V23_A
-0.5	.12	.94	-.3	.94	-.3	0.62	.47	54.5	54.3	V8_A
0.29	.18	.80	-1.5	.79	-1.6	0.62	.55	74.8	70.9	V10_A
0.47	.18	.81	-1.6	.80	-1.6	0.63	.46	71.5	67.7	V36_A
0.44	.17	.80	-1.7	.79	-1.8	0.64	.47	69.1	63.8	V33_A
-0.81	.14	.85	-1.1	.84	-1.2	0.64	.49	61	57.4	V7_A
-0.82	.16	.79	-1.7	.77	-1.8	0.65	.53	71.5	63.7	V25_A
-0.79	.16	.77	-1.7	.74	-1.9	0.66	.49	69.9	66.2	V24_A
0.73	.19	.75	-2.0	.73	-2.1	0.66	.48	74.8	71.4	V31_A
-0.24	.18	.73	-2.2	.70	-2.4	0.69	.44	74.8	70.7	V11_A
0.16	.18	.71	-2.3	.68	-2.5	0.7	.46	76.4	71.2	V16_A
-0.01	.19	.70	-2.4	.66	-2.7	0.71	.45	76.4	72.3	V19_A
0.16	.18	.69	-2.5	.66	-2.7	0.72	.46	76.4	71.2	V17_A
0.51	.18	.66	-3.0	.64	-3.1	0.75	.47	72.4	68.4	V18_A

Figure 1: Item Polarity Value Analysis

Fit Statistics

According to Bond and Fox (2007), the outlier-sensitive statistic (outfit statistic) reveals a large difference between the observed and expected values of an item that is outside the range

of the individual's ability. In contrast, the information-weighted fit statistic (infit statistic) reveals the residual of an item that is within the range of the person's ability. The outfit statistic of Mean Square (MNSQ), according to Planinic et al. (2019), offers a more pronounced calculation than the infit statistics in Rasch analysis. Aside from the MNSQ infit and outfit values, the infit and outfit of Z-Standardized (ZTSD) should be in the -2 to +2 range. However, if the MNSQ value meets item suitability and sample adequacy criteria, the ZSTD value may be ignored (Bond & Fox, 2007; Linacre, 2002). Figure 2 shows the item fit measurement.

Measure	Model S.E.	Infit		Outfit		PT-MEASURE		Exact OBS%	Match EXP%	ITEM
		MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
.49	.35	1.16	.6	1.09	.4	0	.26	92.7	92.7	V29_A
1.75	.19	1.37	3.2	1.43	3.2	0.04	.43	67.5	68.1	V22_A
1.26	.14	2.59	6.4	2.34	5.3	0.07	.62	67.5	66.7	V42_A
1.71	.18	1.38	3.3	1.41	3.4	0.1	.45	68.3	65.6	V20_A
-0.75	.15	1.56	3.2	1.62	3.6	0.14	.49	66.7	64.2	ACTIVITI
2.49	.20	1.14	1.5	1.39	2.7	0.15	.39	69.1	70.3	V34_A
-0.47	.12	1.85	4.9	2.08	5.4	0.17	.55	49.6	56.3	V39_A
-0.92	.25	1.22	1.2	1.30	1.3	0.18	.39	79.7	84.5	SKILLSQ2
0.79	.20	1.11	1.3	1.15	1.3	0.32	.42	67.5	70.3	V13_A
-0.57	.17	1.17	1.3	1.18	1.4	0.33	.47	63.4	64.9	V32_A
-0.92	.25	1.06	.4	1.07	.4	0.34	.39	82.9	84.5	OVERALLC
0.52	.17	1.15	1.2	1.18	1.4	0.35	.48	69.1	66.3	V40_A
-0.94	.16	1.08	.4	1.23	1.1	0.36	.45	74	72.8	SKILLSOT
-0.64	.16	1.11	.7	1.07	.5	0.4	.47	66.7	68.5	SKILLSQ3
1.64	.24	.89	.7	.89	-.5	0.44	.42	82.9	82.2	V14_A
-1.2	.16	1.05	.4	1.05	.4	0.45	.32	65	65.5	LANGUAGE
-0.89	.17	1.02	.2	1.00	.0	0.45	.48	72.4	68.9	V41_A
0.06	.20	.95	-.3	.95	-.2	0.45	.46	80.5	77.1	V46_A
0.33	.18	.92	-.5	.93	-.5	0.47	.43	75.6	71.6	V45_A
-1.08	.29	.77	-1.1	.83	-.5	0.51	.45	91.1	88.4	SUBJECTA
-0.44	.12	1.15	1.1	1.13	.9	0.53	.36	55.3	52.4	LAYOUTAN
-0.88	.17	.91	-.6	.91	-.6	0.54	.56	73.2	67.4	V44_A
0.35	.18	.90	-.7	.90	-.7	0.54	.47	68.3	69.4	V21_A
0.23	.18	.89	-.8	.89	-.8	0.54	.46	74.8	71.1	V43_A
0.31	.21	.84	-1.0	.80	-1.2	0.55	.46	78.9	77.9	V15_A
-0.94	.17	.87	-1.0	.86	-1.0	0.58	.41	69.9	66.2	V35_A
-0.26	.14	.91	-.7	.89	-.8	0.59	.48	56.9	56.4	PRACTICA
0.29	.18	.82	-1.4	.81	-1.4	0.6	.54	74.8	70.9	V37_A
-0.92	.16	.82	-1.0	.81	-1.1	0.61	.46	75.6	66.5	V23_A
-0.5	.12	.94	-.3	.94	-.3	0.62	.47	54.5	54.3	V8_A
0.29	.18	.80	-1.5	.79	-1.6	0.62	.55	74.8	70.9	V10_A
0.47	.18	.81	-1.6	.80	-1.6	0.63	.46	71.5	67.7	V36_A
0.44	.17	.80	-1.7	.79	-1.8	0.64	.47	69.1	63.8	V33_A
-0.81	.14	.85	-1.1	.84	-1.2	0.64	.49	61	57.4	V7_A
-0.82	.16	.79	-1.7	.77	-1.8	0.65	.53	71.5	63.7	V25_A
-0.79	.16	.77	-1.7	.74	-1.9	0.66	.49	69.9	66.2	V24_A
0.73	.19	.75	-2.0	.73	-2.1	0.66	.48	74.8	71.4	V31_A
-0.24	.18	.73	-2.2	.70	-2.4	0.69	.44	74.8	70.7	V11_A
0.16	.18	.71	-2.3	.68	-2.5	0.7	.46	76.4	71.2	V16_A
-0.01	.19	.70	-2.4	.66	-2.7	0.71	.45	76.4	72.3	V19_A
0.16	.18	.69	-2.5	.66	-2.7	0.72	.46	76.4	71.2	V17_A
0.51	.18	.66	-3.0	.64	-3.1	0.75	.47	72.4	68.4	V18_A

Figure 2: Measurement of Item Misfit Order

However, if items fulfill one criterion, the item must be retained (Sumintono & Widhiarso, 2015). As shown in Table 5, all the items met at least one criterion except two (V42_A and ACTIVITI). As a result, only two items were removed from this instrument.

Table 5: Item Misfit Order

Item	Outfit MNSQ (0.50-1.50)	Outfit ZSTD (-2.0-2.0)	PTMEA-CORR (0.40-0.85)	Result
V49 A	1.09	0.4	0	Retained
V22 A	1.43	3.2	0.04	Retained
V42 A	2.34	5.3	0.07	Removed
V20 A	1.41	3.4	0.1	Retained
ACTIVITY	1.62	3.6	0.14	Removed
V34 A	1.39	2.7	0.15	Retained
V39 A	2.08	5.4	0.17	Retained
SKILLSQ2	1.30	1.3	0.18	Retained
V13 A	1.15	1.3	0.32	Retained
V32 A	1.18	1.4	0.33	Retained
OVERALL	1.07	0.4	0.34	Retained
V40 A	1.18	1.4	0.35	Retained
SKILLS	1.23	1.1	0.36	Retained
V31 A	0.73	-2.1	0.66	Retained
V11 A	0.70	-2.4	0.69	Retained
V16 A	0.68	-2.5	0.70	Retained
V19 A	0.66	-2.7	0.71	Retained
V17 A	0.66	-2.7	0.72	Retained
V18 A	0.64	-3.1	0.75	Retained

Dimensionality Analysis

Table 6: Dimensionality Analysis Results

	Empirical			Modeled
Total raw variance in observations	68.7	100.0%		100.0%
Raw variance explained by measures	26.7	38.8%		39.4%
Raw variance explained by persons	10.4	15.2%		15.4%
Raw Variance explained by items	16.2	23.6%		24.0%
Raw unexplained variance (total)	42.0	61.2%	100.0%	60.6%
Unexplained variance in 1st contrast	4.6	6.7%	10.9%	
Unexplained variance in 2nd contrast	4.0	5.8%	9.6%	
Unexplained variance in 3rd contrast	2.8	4.0%	6.6%	
Unexplained variance in 4th contrast	2.5	3.7%	6.1%	
Unexplained variance in 5th contrast	2.1%	3.1%	5.0%	

Based on the dimensionality analysis result, the score of raw variance explained by empirical measures is 38.8%, while the Rasch model predicts 39.4%. In this case, the empirical construct validity has almost the same value as the predictions of the Rasch model. The construct validity results have acceptable criteria because they meet the unidimensionality criteria of higher than 20%. The score of the first to fifth unexplained variance is below 15%, which means the instrument uniformity is in a good category. This indicates that the questions used in this study relate to the material's content (Musa et al., 2017). This is supported by Saidi and Siew's research (2019), which states that the Raw variance explained by measures higher than 20% is acceptable, higher than 40% is good, and higher than 60% is excellent. Besides, unexplained variance for 1 to 5 contrast less than 12%, which falls within the ideal range value of less than 15%. The analysis results reveal that the study's construct validity shows the uniformity of the instruments, which are in the good category. This shows that the questions used in this study are related to the material's content.

Reliability Analysis

The Rasch Model analysis evaluates the reliability and separation of items and persons. This statistic demonstrated the items' ability to separate persons with varying levels of the measured concept.

Reliability result

According to Table 7, a person's reliability higher than 0.81 is interpreted as “good.” A high separation value indicates that the instruments have good quality since they can identify the group of respondents. In this study, the value of Person Reliability is 0.90, with a Person Separation value of 2.97 based on Table 8. So, the instruments have good quality.

Table 7: Reliability in Rasch Analysis

Statistics	Fit Indices	Interpretation
Item and Person Reliability	<0.67	Low
	0.67-0.80	Sufficient
	0.81-0.90	Good
	0.91-0.94	Very Good
	>0.94	Excellent
Item and Person Separation	≤ 2	A high separation value indicates that the instruments are of good quality since they can identify the group of items and respondents.

Source: Sumintono & Widhiarso, 2014

Table 8: Person Reliability and Separation Index

	Raw Score	Count
Mean	172.0	44
S.D	11.8	0.0
Real RMSE	0.33	
ADJ. SD	0.98	
Separation	2.97	
Person reliability	0.90	
Total person input	123	

According to Table 7, an item's reliability higher than 0.94 is interpreted as “excellent.” For an item’s separation, a high separation value indicates that the instruments have good quality since they can identify the group of items. In this study, the value of Item Reliability is 0.95, with an Item Separation value of 4.32 based on Table 9. So, the instruments have good quality.

Table 9: Item Reliability and Separation Index

Mean	480.3	123.0
S.D	37.9	0.0
Real RMSE	0.20	
ADJ. SD	0.85	
Separation	4.32	
Item reliability	0.95	
Total item input	42	

Conclusion

To ensure that the developed instrument can be used again, it must be developed accurately and appropriately in terms of validity and reliability. Instruments that have been correctly developed will have no difficulty measuring the variables being researched (Hassan et al., 2019). This condition will surely help researchers in concluding the analysis of the findings. The findings of Rasch for the developed instrument proved that the newly developed ESL

textbook evaluation questionnaire is valid and reliable as the instrument can be used as a measurable instrument to evaluate ESL textbooks in Malaysian community colleges. Furthermore, the analysis conducted using the Rasch Measurement Model to assess the validity and reliability of the newly developed ESL textbook evaluation questionnaire has connected theoretical research with practical application. The investigation into validity and reliability in this research offers a different option for upcoming scholars in English Language Teaching (ELT) textbook evaluation to explore the reliability and validity of tools created through Item Response Theory (IRT) rather than the Critical Test Theory (CTT) measurement model. The elements of validity and reliability are crucial and should be maintained, especially when crafting a new research tool. Essentially, the assumptions validated in this research that utilized the Rasch Measurement Model were Item Fit, Polarity, Reliability, Separation Index, and Unidimensionality. This paper enables researchers, particularly those in language studies, to cultivate a new perspective on incorporating the Rasch Measurement Model, which has been relatively underexplored in ELT textbook evaluation. Additionally, there are eight diagnostic data analyses possible with the Rasch Measurement Model, which encompasses (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 standardized 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. Therefore, future researchers might explore additional assumptions of the Rasch Measurement Model not covered in this study, tailored to specific study goals and objectives.

Implication

This study proposed a valid and reliable instrument to evaluate the suitability of an ESL textbook. As a result, this study could benefit lecturers and institution administrators, enabling them to assess the effectiveness of an ESL textbook. Based on the results of the analysis, student feedback can assist lecturers in selecting the most appropriate ESL textbook that meets students' needs and improves the educational experience.

Recommendations

There are some suggestions for subsequent studies. While this study focused on community college students, future investigations could explore other higher educational institutions in Malaysia. Additionally, this research solely included students from higher educational institutions in Malaysia. Future studies might benefit from involving students at primary and secondary schools, colleges, or universities in Malaysia. Engaging different populations may introduce new variables for exploration.

References

- Abdel Hakim, B. (2023). *Textbook Evaluation: "My Book of English" The case of the third year at the primary school of Touggourt District* (Doctoral dissertation, Kasdi Merbah Ouargla University).
- AbdelWahab, M. M. (2013). Developing an English language textbook evaluative checklist. *IOSR Journal of Research & Method in Education*, 1(3), 55–70
- Acton, G. S. (2003). What is good about Rasch measurement? *Rasch Measurement Transactions*, 16(4), 902–903.
- Ahmad, M., Ismail, M.K.A., & Saba, N. (2019). Evaluating an English language textbook: A study on 'English-2' in the light of Communicative Language Teaching principles. *Revista Dilemas Contemporáneos: Educación, Política y Valores*, VI, 1–38
- Ahour, T., & Ahmadi, E. (2012). Retrospective evaluation of textbook "Summit 2B" for its suitability for EFL undergraduate students. *Journal of Educational and Social Research*, 2(5), 195–202.
- Ahour, T., Towhidiyan, B., & Saeidi, M. (2014). The evaluation of 'English Textbook 2' taught in Iranian high schools from teachers' perspectives. *English Language Teaching*, 7(3), 150–158. <https://doi.org/10.5539/elt.v7n3p150>
- Akbar, R. (2016). An analysis of selected eleventh-grade English textbooks. *Journal of English and Education*, 4(1), 109–126. Retrieved from <https://media.neliti.com/media/publications/192500-EN-none.pdf>
- Amerian, Majid, & Khaivar, Alimorad. (2014). Textbook selection, evaluation, and adaptation procedures. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, 6 (1), 2289-3245. Retrieved from https://www.researchgate.net/publication/320172366_Textbook_Selection_Evaluation_and_Adaptation_Procedures
- Amin, J. B. (2016). Quality assurance of the qualification process in TVET: Malaysia Country. *The Online Journal for Technical and Vocational Education and Training in Asia*.
- Anderson, L. M. (1989). Learners and learning. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher*. pp. 85–99. Pergamon Press.
- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert review of pharmacoeconomics & outcomes research*, 11(5), 571–585.
- Aswini, M. G., Rao, M. M. S., Sridevi, D., & Pushpalatha, M. J. (2023). The Evolution of Coursebook Evaluation Checklists. *Journal of Advanced Zoology*, 44(S2), pp. 2867–2874.
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Ed. 3). New York: Routledge.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, 15(4), 1–7. <https://doi.org/10.1187/cbe.16-04-0148>
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the Human Sciences*. <https://doi.org/10.1007/978-94-007-6857-4>
- Byrd, P. (2001). Textbooks: Evaluation for Selection and Analysis for Implementation. In M. Celce- Murcia (Eds) *Teaching English as a second or Foreign Language* (3rd ed. Pp 415–415) Us. Heinle
- Cavanagh, R. F., & Waugh, R. F. (Eds.). (2011). *Applications of Rasch measurement in learning environments research* (Vol. 2). Springer Science & Business Media.
- Chang, Y. F. (2002). *A study on English textbook selection for elementary school*. ChungHsing: National ChungHsing University.

- Chinedu, C. C., & Wan-Mohamed, W. A. (2017). Realigning SD Goals for Industry and TVET Training Programs: A Crucial Undertaking. *The Online Journal for Technical and Vocational Education and Training in Asia*
- Combrinck, C. (2020). Big changes in achievement between cohorts: A true reflection of educational improvement, or is the test to blame? *Rasch Measurement: Applications in Quantitative Educational Research*, 179-193.
- Creswell, J. W. (2012). *Research Design: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*.
- Cunningsworth, A. (1995). Choosing your coursebook. Heinemann English Language Teaching
- Deuri, Champak. (2012). An evaluative study of textbooks in English at the higher secondary level. *International Journal of Science, Environment, and Technology*, 1(1). Retrieved from <http://www.ijset.net/journal/15.pdf>
- Dubin, F. & Olshtain, E. (1986). *Course Design*. New York: CUP.
- Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press
- Emelyanova, N., & Voronina, E. (2014). Introducing a learning management system at a Russian university: Students' and teachers' perceptions. *International Review of Research in Open and Distance Learning*, 15(1), 272–289. <https://doi.org/10.19173/irrodl.v15i1.1701>
- Faradillah, A., & Febriani, L. (2021). Mathematical trauma students' junior high school based on grade and gender. *Infinity*, 10(1), 53-68.
- Forkmann, T., Boecker, M., Norra, C., Eberle, N., Kircher, T., Schauerte, P., ... & Wirtz, M. (2009). Development of an item bank for the assessment of depression in persons with mental illnesses and physical illnesses using Rasch analysis. *Rehabilitation Psychology*, 54 (2), 186.
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30.
- Fredriksson, Cecilia, & Olsson, Rebecca. (2006). English textbook evaluation. An investigation into criteria for selecting English textbooks. Student essay 15hp. Retrieved from <http://hdl.handle.net/2043/2842>
- Garinger, D. (2002). Textbook selection for the ESL classroom. *Eric Digest*. Retrieved from http://www.mcael.org/uploads/File/provider_library/Textbook_Eval_CAL.pdf
- Ghezlou, M., Moghtadi, L., & Koosha, M. (2016). EFL teachers' and students' perspectives on Total English Upper Intermediate: An analytical evaluation. *Modern Journal of Language Teaching Methods*, 6(2), 101–116.
- Gholami, R., Noordin, N., & Rafik-Galea, S. (2017). A Thorough Scrutiny of ELT Textbook Evaluations: A Review Inquiry. *International Journal of Education and Literacy Studies*, 5(3), 82-91.
- Gholampour, S., & Mehrabi, D. (2023). Literature Review of ELT Textbook Evaluation. *MEXTESOL Journal*, 47(2), n2.
- Grosch, M. (2017). Developing A Competency Standard for TVET Teacher Education in ASEAN Countries. *Jurnal Pendidikan Teknologi dan Kejuruan*, 279 - 287.
- Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Vol 2. Sage Publications, Incorporated; 1991.
- Hassan, N. F., Puteh, S., Muhamad Sanusi, A., & Che Mohamad Zahid, N. H. (2019). Student perspective on technology enabled/enhanced active learning in education: Rasch measurement model. *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue 3), 929–935. <https://doi.org/10.35940/ijeat.F1061.0986S319>

- Ho, Hsuan-fu, & Hsu, Ying-tsun. (2011). Improving the textbook adoption process in Taiwan. *International Education Studies*, 4, (4), 356-357. Retrieved from www.ccsnet.org/ies
- Huang, S. (2011). *Ideal and reality of textbook selection: an interview- and questionnaire-based investigation in the Taiwanese tertiary context*. The University of Warwick. Retrieved from <http://wrap.warwick.ac.uk/36807/>
- Hutchinson, T. & Torres, E. (1994). The Textbook as Agent of Change. *ELT Journal*, 48(4), 315-328
- Iramaneerat, C., Smith, E.V. & Smith, R.M. (2008). An Introduction to Rasch Measurement. In Osborne, J. (Ed). *Best Practices in Quantitative Methods*. pp. 50–70. Sage Publications.
- Işık, Ali, & Kurum, Eyüp. (2002). How are ELT materials selected in Istanbul? *International 1st Teaching and Learning Symposium*, 1(1), 20-23.
- Johnson, B., & Christensen, L. (2012). Educational Research: Quantitative, Qualitative, and Mixed Approaches 54 www.ccsnet.org/ies *International Education Studies* Vol. 7, No. 11; 2014 (4th ed.). Published by SAGE, India.
- Kay, R. (2006). Evaluating strategies to incorporate technology into pre-service education: A literature review. *Journal of Research on Technology in Education*, 38(4), 383-398
- Kumaravadivelu, B. (2006). TESOL methods: Changing tracks, challenging trends. *TESOL Quarterly*, 40(1), 59-81. <https://doi.org/10.2307/40264511>
- LaBelle, J. (2010). Selecting ELL textbooks: a content analysis of L2 learning strategies. *Journal of Language Teaching and Research*. 1(4), 358–369. doi: https://www.academia.edu/6865104/Marquette_University_Selecting_ELL_Textbooks_A_Content_Analysis_of_L2_Learning_Strategies_Selecting_ELL_Textbooks_A_Content_Analysis_of_L2_Learning_Strategies
- Lemmer, M. (2008). Educators' selection and evaluation of natural sciences textbooks. *South African Journal of Education*, pp. 28, 175–187.
- Linacre, J. M. (2018). Winsteps [Computer program Ver. 3.93. 2]. *Chicago, IL: win steps. com*.
- Linacre, J.M. (2020). Winsteps® (Version 4.5.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2020. Available from <https://www.winsteps.com/>.
- Litz, D.R.A. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal*, pp. 48, 1–53.
- Liu, K., & Diana Deris, F. (2023). Teachers' Perceptions of Communicative Language Teaching Approach in English Grammar Teaching. *Arab World English Journal (AWEJ) Volume, 14*.
- Lord FM, Novick M. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley; 1969.
- Lord FM. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1980.
- Maftukah, S., & Astuti, P. (2021). Strategies employed by EFL teachers at a Vocational High School for selecting appropriate English textbooks. *ELT Forum: Journal of English Language Teaching*, 10(1), 57-69. <https://doi.org/10.15294/elt.v10i1.38795>
- Mahmood, K. (2011). Conformity to quality characteristics of textbooks: the illusion of textbook evaluation in Pakistan. *Journal of Research and Reflection in Education*, 5(2), 170–190. Retrieved from <http://www.ue.edu.pk/jrre>
- Makgato, Moses, & Ramaligela, Sylvia Manto. (2013). Teachers' criteria for selecting textbooks for the technology subject. *African Journal of Research in Mathematics, Science and Technology Education*, 16(1), 32-34. doi: [10.1080/10288457.2012.10740727](https://doi.org/10.1080/10288457.2012.10740727)

- Marczak, M. (2013). Selecting an e-(text) book: Evaluation criteria. *Teaching English with Technology*, 13(1), 29–41. Retrieved from <http://www.tewtjournal.org>
- McDonough, J., Shaw, C., & Masuhara, H. (2017). *Materials and method in ELT: A teacher's guide*. Blackwell Publishing
- Mega, N. D. (2021). *The evaluation of English textbooks based on Litz's theory* (Doctoral dissertation, UIN Raden Intan Lampung).
- Mijayanti, Y. (2015). *SMA English teachers' lived experience in selecting textbooks*. Yogyakarta: Sanata Dharma University
- Monazzah, Z., Behjat, F., & Zareian, A. (2016). General English university coursebook evaluation based on Litz's model and professors' perceptions of such books. *Modern Journal of Language Teaching Methods*, 6(2), 435–442
- Mofreh, SAM, Ghafar, MNA, Omar, AHH, Mosaku, M., & Ma'ruf, A. (2014). Psychometric Properties on Lecturers' Beliefs on Teaching Function: Rasch Model Analysis. *International Educational Studies*, 7 (11), 47-55.
- Mukundan, J. (2011). Developing an English language textbook evaluation checklist: A focus group study. *International Journal of Humanities and Social Science*, 1(12), 105–107. Reprinted from http://www.ijhssnet.com/journals/Vol_1_No_12_September_2011/14.pdf
- Mukundan, J., Hajimohammadi, R., & Nimehchisalem, V. (2011). Developing an English Language Textbook Evaluation Checklist. *Contemporary Issues In Education Research*, 4(6).
- Musa, N. A. C., Mahmud, Z., & Baharun, N. (2017, September). Exploring students' perceived and actual ability to solve statistical problems based on Rasch measurement tools. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012096). IOP Publishing.
- Nourmohammad-Nouri, N., Ghaemi, F., & Sedigh, R. (2015). 'English Result': The qualitative and quantitative analysis of course books perceived by Iranian EFL learners and teachers. *Procedia – Social and Behavioral Sciences*, pp. 192, 601–614. <https://doi.org/10.1016/j.sbspro.2015.06.106>
- Orfan, S. N., Noori, A. Q., & Akramy, S. A. (2021). Afghan EFL instructors' perceptions of English textbooks. *Heliyon*, 7(11).
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 20). Oxford: Oxford University Press.
- Preedy, M. (2001). Curriculum evaluation: Measuring what we value. In D. Middlewood & N. Burton (Eds.), *Managing the Curriculum*. pp. 89–103. Paul Chapman
- Rahimpour, Massoud, & Hashemi, Raheleh. (2011). Textbook selection and evaluation in EFL context. *World Journal of Education*, 1(2), 84-86. Retrieved from www.sciedu.ca/wje
- Ramamurthy, V., Alias, N., & DeWitt, D. (2020). Technical communication in Malaysian skill-based institutions: the need for pedagogical evolution. *Journal of Issue In Education*, 43.
- Rasul, M. S., Ashari, Z. H., Azman, N., & Rauf, R. A. (2015). Transforming TVET in Malaysia: Harmonizing the Governance Structure in a Multiple Stakeholder Setting. *The Online Journal for Technical and Vocational Education and Training in Asia*.
- Reid, E. (2017). How to choose the right English language textbook. *Problemy Wczesnej Edukacji/issues in Early Education*, 1 (36), 1734-1582. Retrieved from http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-4ed97e50-31f9-4ec5-b883-ae847456d112/c/Eva_Reid_Elena_Kova_ikova.pdf

Rodriguez, V. C. (2015). Selecting a science textbook for the CLIL classroom: a checklist proposal. Oviedo: Universidad de Oviedo.

Saidi, S. S., & Siew, N. M. (2019). Reliability and validity analysis of statistical reasoning test survey instrument using the Rasch measurement model. *International Electronic Journal of Mathematics Education*, 14(3), 535–546. <https://doi.org/10.29333/iejme/5755>

Sekaran, U. (2003). *Research Methods for Business: A Skill Building Approach*. Published by John Wiley & Sons, Inc.

Shabani, K. (2017). Evaluation of Iranian second-grade high school English textbook. *Bulletin de la Société Royale de Liège*, 86(2017). 111–126.

Sheldon, E. L. (1988). Evaluating ELT textbooks and materials. *ELT Journal*, 42(4). 237–246. [doi 10.1093/elt/42.4.237](https://doi.org/10.1093/elt/42.4.237)

Sikorova, Z. (2004). *The textbook selection in primary and secondary school*. Oostrava: University of Oostrava.

Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia. *Teaching English as a second or foreign language*. pp. 432–453. (2nd Ed.). Boston: Heinle & Heinle Publishers

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Penerbit Trim Komunikata

Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., ... Zappia, A. (2019). Development and validation of a university student's progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis. *International Journal of Science Education*, 41(3), 388–417. <https://doi.org/10.1080/09500693.2018.1556414>

Ur, P. (1996). *A course in language teaching: Practice and theory*: Ernst Klett Sprachen.

Van, N. T. (2011). Language learners' and teachers' perceptions of learner autonomy- Are they ready for autonomous language learning? *VNU Journal of Foreign Studies*, 27(1), 41–52. <https://js.vnu.edu.vn/FS/article/view/1463>

Watt, M. G. (2009). Research on the textbook selection process in the United States of America. *IARTEM e-Journal*, 2:1. Retrieved from <https://files.eric.ed.gov/fulltext/ED506523.pdf>

Weiss, M. L. (2014). *Routledge Handbook of Contemporary Malaysia*. New York: Routledge.

Wijayanto, P. W. (2017). Industry-Based Curriculum for Vocational School Students. *Asia Pacific Education Conference* ISBN 978-602-14377-0-4.

Winne, P. H., & Marx, R. W. (1982). Students' and teachers' views of thinking processes for classroom learning. *The Elementary School Journal*, 82(5), 493–518. <https://psycnet.apa.org/doi/10.1086/461284>

Wolfe, E. W., Ray, L. M., & Harris, D. C. (2009). *Educational and Psychological Measurement*.

Yuen, K.K. F., & Ting, T. O. (2012). Textbook selection using fuzzy Prometheus ii method. *International Journal of Future Computer and Communication*, 1(1). 23–25. Retrieved from <http://www.ijfcc.org/papers/20-E10008.pdf>