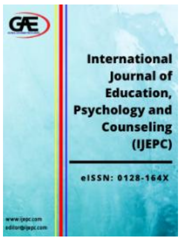
International Journal of
Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC

Volume 11 Issue 64 (September 2026) PP. 1084-1109



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
https://gaexcellence.com/ijepc

GLOBAL ACADEMIC EXCELLENCE

RELIABILITY AND VALIDITY OF THE MALAYSIAN POLYTECHNIC LECTURERS' ASSESSMENT LITERACY TEST INSTRUMENT (UL3PM) USING THE RASCH MODEL: A PILOT STUDY

Normazita Mat Ali[1,4], Razif Salim[2,3] , Nurfirdawati Muhamad Hanafi[4]

[1]Department of Petrochemical Engineering, Politeknik Tun Syed Nasir Syed Ismail, Malaysia
normazita@ptsn.edu.my https://orcid.org/0009-0003-3835-8833
[2]Department of Planning & Management Source, Perpustakaan Tunku Tun Aminah University Tun Hussein Onn Malaysia (UTHM), Malaysia
razifs@uthm.edu.my https://orcid.org/0009-0008-6946-3974
[3]Bank Rakyat School of Business, Innovation, Technology and Entrepreneurshipt, University Tun Abdul Razak (UNIRAZAK)
[4]Faculty of Technical and Vocational Education , University Tun Hussein Onn Malaysia (UTHM),Malaysia
nurfirda@uthm.edu.my https://orcid.org/0000-0001-5571-4383
[*]Corresponding Author

Article Info:

Article history:
Received date: 27.07.2026
Revised date: 20.08.2026
Accepted date: 03.09.2026
Published date: 20.09.2026

Abstract:

This study investigated the reliability and validity of the Malaysian Polytechnic Lecturer Assessment Literacy Test (UL3PM) instrument. The pilot study involved 33 lecturers from Polytechnic Tun Syed Nasir Syed Ismail (PTSN). The instrument contains two sections of questions. Section A includes demographic information, and Part B has 60 multiple-choice questions related to assessment literacy that encompass six dimensions: Assessment Basics; Assessment Planning; Assessment Development; Assessment Implementation, Assessment and Assessment Improvement; and Equity and Ethics in Assessment implemented in Malaysian polytechnics. The data obtained were analysed using the Rasch Model to assess the reliability and validity of the Malaysian Polytechnic Lecturer Assessment Literacy Test (UL3PM) instrument. Overall, it was found that the UL3PM instrument has high item and individual reliability with values of 0.87 and 0.91, respectively. The validity of the UL3PM instrument was investigated

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

through item fit, individual fit, variable maps, and unidimensionality. The study found that the UL3PM instrument is reliable and valid for measuring assessment literacy among Malaysian polytechnic lecturers with minor adjustments to misfitting and redundant items.

To cite this document:

Ali, N. M., Salim, R., & Hanafi, N. M. (2026). Reliability And Validity of The Malaysian Polytechnic Lecturers' Assessment Literacy Test Instrument (Ul3pm) Using the Rasch Model: A Pilot Study. *International Journal of Education, Psychology and Counselling, 11*(64), 1084-1109.

DOI: 10.35631/IJEPC.1164057

Keyword:

Assessment Literacy, Rasch Model, Reliability, UL3PM, Validity

Introduction

In Malaysia's Technical and Vocational Education and Training (TVET) system, lecturers' roles extend beyond content delivery (Giraldo, 2021). They are also responsible for assessing students' practical and cognitive skills capabilities through fair, reliable, and valid assessment approaches (Afshar & Ranjbar, 2021). This responsibility aligns with the broader national education goals, particularly those centred around Sustainable Development Goal 4 (SDG 4), which aims to ensure inclusive and equitable quality education for all (Gan & Lam, 2022; Pastore, 2023). Thus, assessing competent educators in polytechnic institutions is becoming increasingly critical (Aria, Sukyadi, & Kurniawan, 2021; Giraldo, 2021a).

Although assessment literacy (AL) is widely recognised at the global level, there is still a lack of instruments specifically developed to assess this literacy in the context of polytechnics in Malaysia (Tajeddin, Khatib, & Mahdavi, 2022). Most existing instruments are developed within the context of Western Education or designed for the general education system, making them less suitable for application in the TVET environment (Bøhn & Tsagari, 2021). Polytechnic lecturers usually assess practical projects, workplace simulations, and technical assignments that require a specific assessment framework (Fitriyah et al., 2022; Rasul et al., 209). This emphasises the need for an instrument that assesses theoretical knowledge and measures the ability to conduct assessments ethically (Yan & Pastore, 2022; Jean de Dieu & Zhou, 2021).

Although this issue is gaining more attention, studies show that many lecturers in Malaysia still demonstrate limited competencies in foundational assessment practices such as item construction, marking consistency, delivering effective formative feedback, and interpreting assessment data for instructional decisions (Nur et al., 2025; Rozali et al., 2022). The lack of standardised, reliable instruments to evaluate these skills impedes targeted professional development and institutional quality monitoring (Purnami, Hartinah, & Susongko, 2024; Anggara & Abdillah, 2023).

The Polytechnic Malaysia Lecturer Assessment Literacy Test Instrument (UL3PM) was developed to address this issue. This instrument aims to assess lecturers' assessment literacy across key domains such as assessment design, ethics, implementation, and assessment data. The Rasch Measurement Model was employed to ensure the psychometric robustness of this instrument. This model allows assessment of item appropriateness, respondent reliability, item difficulty level, and the competency of the measured construct dimensions (Khalid et al., 2023; Rohaya, 2009). This paper presents findings from the UL3PM pilot test and the Rasch analysis and evaluates its suitability for wider use.

Literature Review

Assessment Literacy in TVET Context

Assessment literacy refers to educators' understanding and ability to design, implement, interpret and utilize assessment results to improve teaching and learning (Njuki & Mukundi, 2023; Singh et al., 2023). In the context of Technical and Vocational Education and Training (TVET), these competencies are even more vital, as they assess cognitive, psychomotor, and affective domains aligned with learning outcomes and industry demands (Diao & Hu, 2022; Vinodhen, 2025). Studies show that many polytechnic lecturers in Malaysia face challenges in developing valid test items, ensuring consistent marking, providing actionable formative feedback, and utilising data for informed instructional decision-making (Mutinda, 2023; Yusop et al., 2023). These constraints are linked to moderate assessment literacy levels and the lack of tailored tools and training (Hashim et al., 2021; Diao, Han, Zhou, & Wang, 2023).

Assessment literacy in TVET should combine theoretical understanding of assessment types (e.g., formative and summative) with practical skills such as developing rubrics, analysing assignments, and formulating performance criteria suited to technical competencies (Asres & Abebe, 2024). These skills are crucial for measuring complex learning outcomes in authentic environments using workplace simulations, practical assignments, and portfolios (Ramaligela, 2022; Yusop et al., 2022). Ethical grading, transparent communication of criteria, and timely, constructive feedback are also essential as these support student motivation, self-regulated learning, and lifelong learning trajectories (Jaya & Nasir, 2025).

Malaysia's efforts to transform the TVET system, based on the national education development plan and global frameworks such as Sustainable Development Goal 4 (SDG 4), also support the need for high-quality assessment literacy (Yusop et al., 2023; Yusop et al., 2022). Strengthening assessment literacy among TVET educators is crucial to ensure equity in educational delivery and support students' transition to the workforce (Diao & Hu, 2022). Therefore, polytechnic lecturers' ability to accurately interpret student performance and use assessment data to make teaching decisions is a key component of a responsive and quality-

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109
assured TVET system (Diao et al., 2023; Ismail et al., 2022). Most of the studies were conducted in English-speaking countries (e.g., the US, Canada, Australia, and New Zealand), while only three studies were located in the European area and some from Malaysia (Njuki & Mukundi, 2023; Hashim et al., 2021; Jaya & Nasir, 2025)

After researching several assessment literacy instruments both domestically and internationally, it was found that no instrument was suitable for measuring the assessment literacy levels of polytechnic lecturers. Although all these instruments are based on STCEAS, several features make them unsuitable for measuring the assessment literacy levels of polytechnic lecturers. The first characteristic is that items in foreign instruments, such as the Assessment Literacy Inventory (ALI) by Mertler and Campbell (2005), use classroom scenarios from Western countries that differ from Malaysian classroom scenarios, especially at polytechnics. This is because the main purpose of constructing the instrument is to test knowledge and skills in assessment. Therefore, the instrument is not suitable for measuring the level of assessment literacy among polytechnic lecturers because the Malaysian polytechnic curricula and scenario are different from Western countries. Indeed, the obstacles confronting each individual from various countries, institutions, and levels of Education are unique.

The Importance of Validity and Reliability of Instruments

Validity and reliability of instruments are important to ensure that measurements of constructs such as assessment literacy are accurate (Modarresi, Jalilzadeh, Coombe, & Nooshabadi, 2021). Validity ensures that the instrument measures what it is intended to measure, while reliability refers to the consistency of results across different times and contexts (Sun & Zhang, 2022; Wright & Master, 1982). Reliability enhances repeatability and reduces measurement error, while validity safeguards the accuracy and relevance of inferences drawn from the results (Fitriyah, Masitoh, & Widiati, 2022). In Malaysia, the lack of locally developed and psychometrically validated instruments often leads to inconsistencies between assessment practices and intended learning outcomes (Padmadewi, Artini, Ratminingsih, & Marsakawati, 2022).

Appropriate and precise data collection instruments are crucial for ensuring that quality, relevant, and accurate data and information are obtained (Mudau, 2022). Thus, the results from processing and analyzing data to answer and achieve the outlined objectives can be accepted as valid and reliable (Fauzan, Arifin, Lubis, & Firdaus, 2022).

Reliability refers to the consistency of an instrument or measuring tool in achieving similar results when used at different times (Hannigan, Alonzo, & Oo, 2022). The reliability of an instrument or measurement tool refers to the consistency of the measurements it produces (Tsagari, 2021). An instrument or measuring tool with a high degree of consistency is considered good and suitable when the obtained scores are nearly the same (Oo, Alonzo, & Asih, 2022).

Meanwhile, the validity of an instrument refers to the extent to which it measures what it is supposed to measure (Kao, 2023). An instrument is valid and accurate if it measures what it is supposed to (Lutovac & Flores, 2022).

Recent literature emphasises that the reliability of an assessment depends critically on the data collection instrument (Bøhn & Tsagari, 2021). Selecting appropriate tools is fundamental, as it directly affects measurement accuracy and reliability (Istiara & Hastomo, 2023). Reliability ensures repeatable results and minimises measurement error, while validity confirms that the instrument accurately assesses the intended construct (Mardiana, 2024). Inaccurate or non-validated instruments risk misleading conclusions (Mudau, 2022). In Malaysia, the scarcity of locally developed, psychometrically validated instruments often lead to inconsistencies between assessment outcomes and intended learning objectives.

Model Rasch as a Psychometric Approach

Reliability indices are essential for evaluating the consistency and precision of measurement instruments (Rasch, 1980). A good instrument should demonstrate unidimensionality, a logical item hierarchy, and acceptable fit statistics, such as MNSQ values between 0.6 and 1.4 (Linacre, 1994; Istiara & Hastomo, 2023).

The Rasch model is gaining increasing attention in educational measurement because of its strong ability to validate instruments across various learning contexts (Rasch, 1980; Mardiana, 2024). The Rasch model enables researchers to assess internal consistency, item appropriateness, and unidimensionality, key indicators of psychometric quality (Bond & Fox, 2007; Tsagari, 2021). Systematic reviews have shown that Rasch modelling can significantly enhance scale accuracy, diagnostic capability, and instrument calibration (Bøhn & Tsagari, 2021; Mudau, 2022; Arjana et al., 2025).

Several studies in Malaysia have applied the Rasch model to evaluate assessment instruments in higher Education (Padmadewi et al., 2022; Chong et al., 2021). For example, researchers have used the Rasch measurement model to examine the validity and reliability of authentic learning instruments in Malaysian polytechnics, identifying items that were well targeted to respondents and others that overlapped in difficulty (Fauzan et al., 2022).

Similarly, Rasch analysis has been applied to assess the psychometric quality of a computational thinking instrument among secondary school students, revealing both well-functioning items and redundancy in certain difficulty levels (Boone, 2021; Fitriyah et al., 2022). These findings align with broader evidence that while many test items are appropriate, overlap in difficulty can reduce an assessment's discriminatory power (Sun & Zhang, 2022; Christoforidou & Kyriakides, 2021; Linacre, 2017; Muhammad et al., 2020).

The Rasch approach has also been used to evaluate the Academic Enhancement Support instrument for Malaysian student-athletes (Oo et al., 2022; Brown & Abeywickrama, 2019). The study involved respondents from various higher education institutions (Modarresi et al., 2021; Hamzah et al., 2022; Mokshein et al., 2022). Researchers used the Rasch model to examine the instrument's validity and reliability, identifying overlapping items that reduced measurement precision (Lutovac & Flores, 2022; Norhasim & Mohamad, 2020). These redundant items were removed, resulting in a refined instrument with improved psychometric quality (Kao, 2023; Herppich et al., 2018).

In addition, the reliability of the Attitude and Perception towards Academic and Character instrument has been tested using several analyses under the Rasch measurement model

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 64 (September 2026) PP. 1084-1109
(Hannigan et al., 2022). Four items were found to be less suitable and needed to be removed (Rozali et al., 2022; Osman et al., 2022). As a result of these improvements, the instrument became more appropriate and accurate because its validity and reliability were enhanced (Mardiana, 2024; Ismail et al., 2022).

Several psychometric criteria are needed to determine the validity of a constructed instrument (Mudau, 2022). First, instrument items must be capable of measuring both the respondents and the construct items developed, and differentiating the study respondents (Tsagari, 2021). Second, the items must be suitable for the assessment being measured (Istiara & Hastomo, 2023). Constructed items should fall within the desired logit range for instrument studies using the Likert scale, with item values ranging from 0.6 to less than 1.4 (Mardiana, 2024). Third, having a rating scale without disturbances is critical (Fitriyah et al., 2022). An instrument free from disturbances helps respondents choose their desired answers, ensuring an equal probability for an individual to select any category (Sun & Zhang, 2022). Fourth, unidimensional validity must be ensured so that the measurement is unidimensional (Kao, 2023; Juanjuan et al., 2022).

The instrument's reliability is determined using item and individual reliability values, as well as item and individual separation indices (Wright & Mok, 2004; Rozali et al., 2022). A reliability value exceeding 0.8 is considered strong and acceptable, while a value between 0.6 and 0.8 is less acceptable, and a value below 0.6 is unacceptable (Oo et al., 2022; Jonsson & Svingby, 2007). The item separation index indicates the separation for the item's difficulty level, while the individual separation index shows the number of ability strata within the respondent group (Modarresi et al., 2021; Khalid et al., 2021). A discrimination value greater than 2 is considered good and acceptable because it indicates that the items in this construct can differentiate respondents into at least two ability levels (Lutovac & Flores, 2022; Khalid et al., 2023).

The instrument's validity is determined through point correlation values, item polarity, and item matching (Fauzan et al., 2022; Khine, 2020). The polarity of an item or item consistency indicates that the items used move in the direction intended by the measured construct (Padmadewi et al., 2022). The initial detection of construct validity refers to the item polarity or point measure correlation value (Hannigan et al., 2022). Positive correlation values for all items indicate that the items function consistently to measure the intended construct (Bøhn & Tsagari, 2021; Kalajahi & Abdullah, 2016). Meanwhile, negative item correlations must be reviewed to determine whether items should be improved or discarded (Kao, 2023; Linacre, 2007). This is because such items may reflect inconsistencies in how individuals respond to the variable being measured (Mudau, 2022; Linacre, 2011)

The item-person map shows whether an instrument is suitable or aligned with the respondent's ability (Rozali et al., 2022; Linacre, 2002). On the left side of the dashed line of the item-person map is the position of the individual or respondent, while the right side is the position of the item (Mudau, 2022; Linacre, 2005). The upper left part of the map shows the position of individuals with high ability, and the upper right part shows the position of items with high difficulty (Padmadewi et al., 2022; Linacre, 2003). The item fit in the Rasch Model measures a latent variable (Tsagari, 2021; Ahn & Bong, 2020). The acceptable range for the mean square statistic (MNSQ) for dichotomous data is 0.7 to 1.3 (Istiara & Hastomo, 2023). For polytomous data, the acceptable range is between 0.6 to 1.4 (Mardiana, 2024). The fit can also be assessed through the infit MNSQ and outfit MNSQ values (Sun & Zhang, 2022). These values should fall within the range of 0.6 to 1.4 (Fauzan et al., 2022). When evaluating item fit, prioritize the outfit MNSQ value over the infit MNSQ value (Kao, 2023). If the items fall outside that range, they need to be isolated so that adjustments can be made (Lutovac & Flores, 2022). This is because item suitability affects the instrument's reliability and validity (Modarresi et al., 2021; Atjonen et al., 2022; Linacre, 2010).

UL3PM Instrument Development

The Malaysian Polytechnic Lecturer Assessment Literacy Test Instrument (UL3PM) was developed to assess assessment literacy among polytechnic lecturers in Malaysia. It encompasses six primary constructs: Basic Concepts of Assessment, Assessment Planning, Assessment Development, Assessment Implementation, Assessment Feedback and Improvement, and Equity and Ethics. Each construct is further divided into sub-constructs to enable comprehensive measurement across knowledge and application domains. Using multiple-choice items along with Rasch scale scoring allows the assessment to be conducted objectively and with high psychometric quality. Pilot testing and validation processes using the Rasch model confirm this instrument's reliability and construct validity, making it suitable for diagnostic and professional development purposes in the TVET environment.

According to Siti Khalid et al. (2021), developing assessment literacy instruments tailored to the local context ensures that the measurements conducted align with national education priorities and local teaching culture. This requirement is particularly important in Malaysian TVET, where teaching responsibilities often differ from those in mainstream academic institutions. Similarly, Modarresi et al. (2021) emphasise using robust validation methods such as Rasch analysis to ensure construct clarity and response consistency. These studies reinforce the rationale for constructing the UL3PM using an evidence-based psychometric approach.

The Methodology

Data Collection Instruments

In this pilot study, the researchers have employed probability sampling. Sampling is intended to be applied to selected individuals because they have experiences at the centre of the phenomenon (Creswell, 2009). Probability sampling techniques use some form of random selection and allow calculation of sampling error, thereby reducing selection bias. Thus, 33 polytechnic lecturers from PTSN were chosen for this pilot study, representing a variety of backgrounds, and they were surveyed, comprising 13 males (39%) and 20 females (61%). They were selected using simple probability. To proceed with the data collection procedure, authorisation from the Research and Innovation Division, Department of Polytechnic and Community College Education, Ministry of Higher Education Malaysia is required. Henceforth, the researcher must obtain permission from the Director of Polytechnic Tun Syed Nasir Syed Ismail (PTSN) before meeting with the respondents. Respondents were invited to participate in the research until the required sample size was attained.

Literacy Test Instrument for the Assessment of Malaysian Polytechnic Lecturers (UL3PM)
The research instrument used to collect data is a test consisting of four multiple-choice questions from the Malaysian Polytechnic Lecturer Assessment Literacy Test (UL3PM). The

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 64 (September 2026) PP. 1084-1109

test includes 60 items (questions) to measure Malaysian polytechnic lecturers' assessment literacy. The Malaysian Polytechnic Lecturer Assessment Literacy Test (UL3PM) consists of two parts, referred to as Part A and Part B.

Part A contains several questions on the respondents' general demographics, as well as their involvement in assessment literacy-related activities. Part B contains sixty (60) questions about the respondents' understanding of the Basics of Assessment, including Assessment Planning, Development, Implementation, Analysis, Improvement, and Equity and Ethics in Assessment. The answers in Part B use a dichotomous scale, providing respondents with four answer choices: A, B, C, or D. The number of correct answers reflects the level of assessment literacy among polytechnic lecturers.

The Basic Assessment Construct has five sub-constructs: the Concept of Assessment, the OBE Concept, the Importance of Assessment, the Principles of Assessment, and the Purpose of Assessment. Table 1 displays all 17 items or questions asked under the Basic Assessment Construct.

Table 1: List of items for the Basic Assessment Constructs

Sub -construct	Attribute	Item
Assessment concept	Concepts of Assessment, testing, measurement and evaluation	S2,S3,S1
	The concept of formative and summative Assessment	
	The concept of competency-based Assessment	S11
		S57
OBE concept	Knowledge and understanding of the principle of Outcome Based Education (OBE)	S10
	Ability to align learning outcomes, teaching and learning activities, and assessment methods.	
	Understanding of constructive alignment in the implementation OBE	
		S12
		S14
Importance of Assessment	Understanding of the importance of assessment in measuring students' achievement of learning outcome	S6

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

Principles of Assessment	Knowledge and understanding of the fundamental principles of quality assessment Understanding of validity, reliability, fairness, transparency and practicality in assessment. Ability to apply appropriate assessment principles to ensure accurate and meaningful assessment result	S8 S18 S4
Purpose of Assessment	Understanding of the purpose and functions of assessment in the teaching and learning process. Ability to distinguish assessment for learning, of learning, and as learning Understanding of the use of assessment information for feedback, grading, decision-making and improvement of student learning	S9,S5 S32 S33

Table 2 present the items under Assessment Planning Construct. This construct comprises five sub-constructs under the Assessment Planning Construct; Assessment Alignment, Assessment Methods, Assessment Specification and Assessment Criteria and Standard, and Student involvement in Assessment. A total of seven items were developed to represent the Assessment Planning Construct.

Table 2: List of items for the Assessment Planning Construct

Sub-construct	Attribute	Items
Assessment Alignment	Alignment of assessment with Course Learning Outcomes (CLOs). Alignment between CLO, teaching and learning activities and assessment Alignment with appropriate PLO	S22 S17 S15
Assessment Methods	Selection of appropriate assessment methods according to learning outcomes.	S19

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

Assessment Specifications	Understanding and preparation of JSU Distribution of items according to cognitive level and learning outcome	S16
Assessment Criteria and Standard	Selection of suitable scoring approaches, marking schemes or rubrics.	S25
Student involvement in Assessment	Planning opportunities for self-assessment and peer assessment.	S59

Referring to Table 3, the Assessment Development Construct has three sub-constructs: Assessment Item Development, Assessment Review Procedure, and Assessment Rubric Development. This construct has 10 items or questions.

Table 3: List of items for the Assessment Development Construct

Sub-construct	Attribute	Items
Assessment Item Development	• Development of objective assessment items	S36
	• Development of subjective response items	S39
	• Development of performance-based assessment task	S24
Assessment Review procedure	• Review of assessment items	S31
	• Review of marking schemes	S34
	• Checking alignment with CLOs and assessment specifications	S35
Assessment of Rubric Development	• Development of scoring criteria	S23
	• Development of holistic rubrics	S60
	• Development of analytics rubrics	S37
	• Establishment of performance levels and descriptors	S38

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

The Assessment Implementation construct has three sub-constructs: Assessment Administration, Scoring and Grading, and Assessment Record and Reporting. As shown in Table 4, this construct has six items or questions.

Table 4: List of items for the Assessment Implementation Construct

Sub-construct	Attribute	Items
Assessment Administration	Implementation of appropriate assessment procedures	S26,S28,S30,S27
Scoring and Grading	Application of appropriate scoring and grading procedures	S29
Assessment Record and Reporting	Recording and reporting of assessment result	S41

Referring to Table 5, which pertains to the Feedback and Improvement Construct. There are 13 items in the sub-constructs of Assessment Feedback, Student Performance Analysis, Assessment Quality Analysis, Follow-up and Remedial Action, and Assessment Improvement.

Table 5: List of items for the Assessment Feedback and Improvement Constructs

Sub-construct	Attribute	Items
Assessment Feedback	Provision of effective and constructive feedback	S42,S53
Student Performance Analysis	Analysis and interpretation of assessment results	S48,S45,S47,S46,S44
Assessment Quality Analysis	Evaluation of quality of assessment items and instruments	S50
Follow up and Remedial Action	Use of assessment results for appropriate intervention	S52,S43,S51
Assessment Improvement	Continuous improvement of assessment practices	S7,S31

Table 6 lists items in the Equity and Ethics Construct in the Assessment. This construct has five sub-constructs: Assessment Fairness, Bias-Free and Inclusive Assessment, Transparency in Assessment, Academic Integrity, and Confidentiality Assessment Ethics. This construct has six items.

Table 6: List of items for the Equity and Ethics Constructs

Sub-construct	Attribute	Items
Assessment Fairness	Fair and Equitable assessment practices	S55
Bias-Free and Inclusive Assessment	Recognition and prevention of assessment bias	S20

International Journal of
Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 64 (September 2026) PP. 1084-1109

Transparency in Assessment	Clear communication of assessment requirement	S21
Academic Integrity	Maintenance of integrity throughout the assessment process	S40
Confidentiality and Assessment Ethics	Ethical management of assessment information and materials	S56,S54

Results And Discussion

Respondent profile from Pilot survey

Primary data were obtained through face-to-face pilot surveys with 33 lecturers at Politeknik Tun Syed Nasir (PTSN). Three departments in PTSN were involved in the pilot survey. The respondents' profile, including gender, age, number of respondents by department, highest educational attainment, and work experience, is summarised in Fig. 7, Fig. 8, Fig. 9, Fig. 10, and Fig. 11.

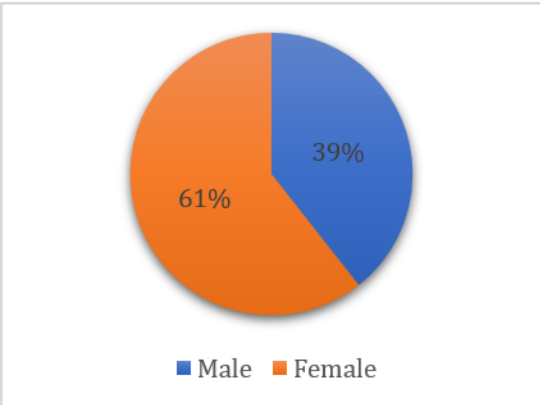


Figure 7: Respondents' Gender

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

Figure 8: Respondents' Ag

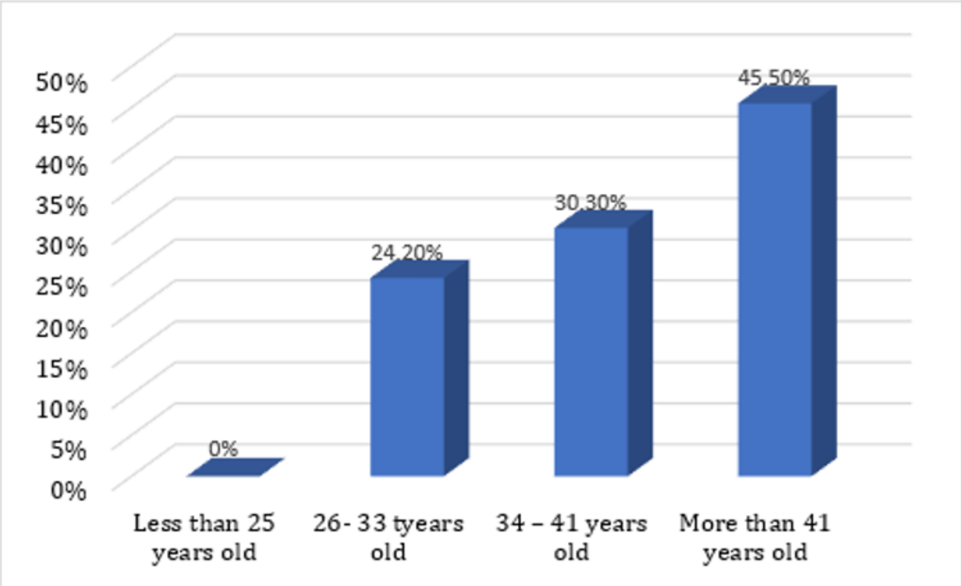


Fig. 8 *Respondents' age*

Figure 8: Respondents' Age

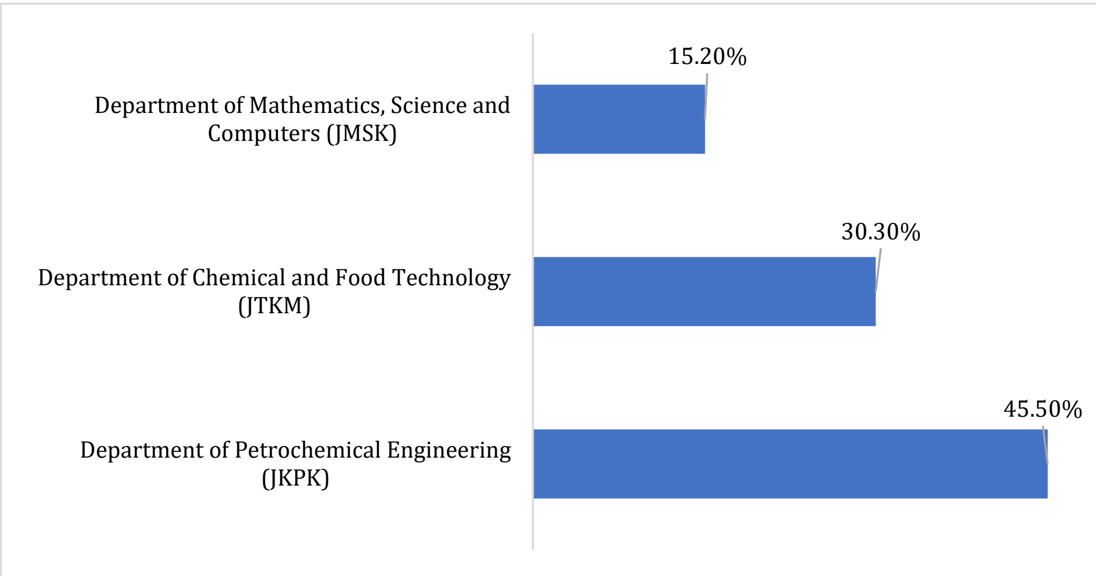


Figure 9: Number of Respondents by Department

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 64 (September 2026) PP. 1084-1109

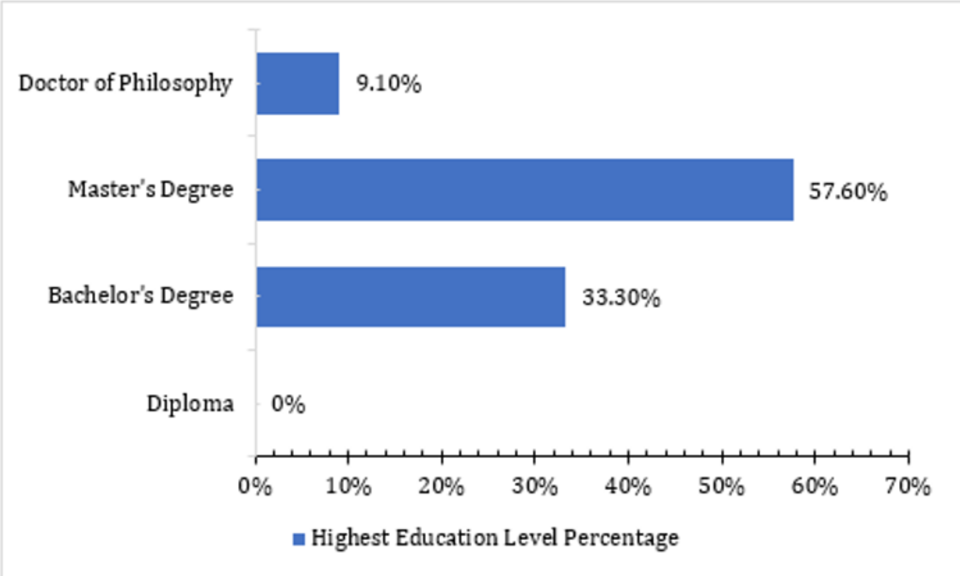


Fig. 10 *The highest level of education*

Figure 10: The Highest Level of Education

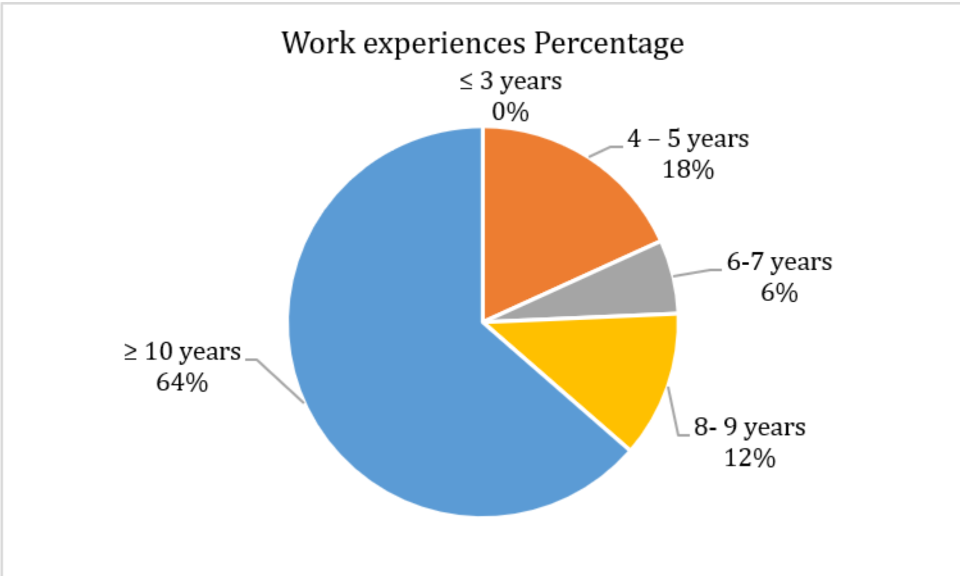


Figure 11: Work Experience Percentage

Figure 7 shows the percentage of male and female lecturers who responded to the UL3PM instrument. Female lecturers who responded to the UL3PM instrument constituted 60.6%, while male lecturers made up 39.4%. The majority of respondents were over 41 years old, accounting for 45.5%, as shown in Figure 8. Figure 9 shows that the majority of respondents are lecturers serving in the Department of Petrochemical Engineering (JKPK). Figure 10 shows that the respondents with a Master's degree who answered the UL3PM instrument the most

International Journal of Education, Psychology and Counseling
EISSN : 0128-164X
IJEPC
Volume 11 Issue 64 (September 2026) PP. 1084-1109

were 57.6%. Figure 11 shows that most respondents have more than 10 years of work experience (63.6%).

The respondents' involvement in assessment-related activities or programs was also inquired about and summarised in Table 7. It was found that 93.9% of individuals had participated in assessment-related programs.

Table 7: Profile follows assessment course/workshop

Attending assessment courses/workshops	Percentage
Yes	93.9%
None	6.1%

Reliability of Ujian Literasi Pentaksiran Pensyarah Politeknik Malaysia (UL3PM) Instrument

Pilot data on 33 polytechnic lecturers serving at Politeknik Tun Syed Nasir Syed Ismail were used to determine the validity and reliability of the UL3PM instrument as an accurate tool in measuring the level of assessment literacy among Malaysian polytechnic lecturers.

The reliability of an instrument refers to the consistency of a measurement tool in assessing the intended construct. In Rasch Model analysis, reliability is measured through several key indicators, namely Person Reliability and Item Reliability, Person Separation Index and Item Separation Index, as well as Cronbach's Alpha (KR-20) for raw scores. Based on Fig.12, the results of the pilot test involving 33 lecturers and 60 items indicate that the Person Reliability is 0.91, while the Item Reliability is 0.87. This indicates that the instrument has a high level of reliability in distinguishing lecturers' ability levels and the clarity of item difficulty levels. The Person Separation value is 3.11, which allows lecturers to be classified into at least 4 different ability levels. This is supported by the strata formula (Linacre, 2002; Linacre,2012):

$$Strata = \frac{4G+1}{3} \approx \frac{4(3.11)+1}{3} \approx 4.15$$

Meanwhile, the Item Separation Index = 2.64, indicating that the number of respondents is sufficient to distinguish items well according to difficulty levels. Additionally, the Cronbach's Alpha (KR-20) value for the raw score is 0.91, indicating very good internal consistency, exceeding the minimum standard of 0.70 recommended in psychometric studies (Bond & Fox, 2015). These results confirm that the Malaysian Polytechnic Lecturer Assessment Literacy Test (ULPPPM) instrument is highly reliable for use in actual studies. This high level of reliability allows researchers to make valid interpretations of the data and to distinguish significantly between individual abilities and item difficulty.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

```
TABLE Analisis Rintis
INPUT: 33 Person  60 Item  MEASURED: 33 Person  60 Item  2 CATS WINSTEPS 3.69.1.11
--------------------------------------------------------------------------------

     SUMMARY OF 33 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      37.3      60.0        .82     .35       1.00     .0    .94    -.1 |
| S.D.      10.1        .0       1.22     .09        .21    1.4    .37    1.1 |
| MAX.      58.0      60.0       4.18     .74       1.67    3.8   2.43    3.4 |
| MIN.      19.0      60.0      -1.08     .31        .60   -3.1    .29   -2.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .37 TRUE SD    1.16  SEPARATION  3.11  Person RELIABILITY  .91 |
|MODEL RMSE    .36 TRUE SD    1.16  SEPARATION  3.23  Person RELIABILITY  .91 |
| S.E. OF Person MEAN = .22                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .91

SUMMARY OF 60 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      20.5      33.0        .00     .47        .99     .1    .94     .1 |
| S.D.       7.4        .0       1.38     .10        .18     .8    .40     .8 |
| MAX.      31.0      33.0       3.00     .75       1.54    2.3   2.27    2.4 |
| MIN.       5.0      33.0      -2.40     .39        .64   -1.7    .29   -1.5 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .49 TRUE SD    1.29  SEPARATION  2.64  Item    RELIABILITY  .87 |
|MODEL RMSE    .48 TRUE SD    1.30  SEPARATION  2.71  Item    RELIABILITY  .88 |
| S.E. OF Item MEAN = .18                                                     |
-------------------------------------------------------------------------------
```

Figure 12: Individual and Item Reliability of the UL3PM

Validity of Malaysia Polytechnic Lecturers' Assessment Literacy Test Instrument Using the Rasch Model

Figure 13 displays the results of the calibration of UL3PM items. It shows the point-measure correlation values, fit statistics, and measure values for 60 items. It shows the extent of construct validity and the quality of the relationship between the items and the respondents. The arrangement of measures needs to be reviewed at this stage. The columns from left to right list the number, raw scores, number of responses, item calibration measure, calibration error, and five fit statistic columns. The PT-Measure Corr column provides correlation reading between person-measure and person-response for each item. The polarity of the items is indicated by the PT-Measure Corr, which shows the extent to which the items function in the same direction and align with the construct being measured. If all values are positive, then it indicates there is no polarity issue. The correlation value will be higher in the middle Part Because there is more variance. However, if the PT-Measure Corr is negative, it indicates that the coding for the respective items does not function in the intended direction.

```
TABLE Analisis Rintis
INPUT: 33 Person  60 Item  MEASURED: 33 Person  60 Item  2 CATS WINSTEPS 3.69.1.11
--------------------------------------------------------------------------------
Person: REAL SEP.: 3.11  REL.: .91 ... Item: REAL SEP.: 2.64  REL.: .87

         Item STATISTICS:  MISFIT ORDER

------------------------------------------------------------------------------------------
|ENTRY   TOTAL                   MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT  MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|     5      6     33     2.71    .52|1.17    .6|2.27   1.9|A .27   .48| 84.8  85.2| S05  |
|    12     29     33    -1.59    .56|1.11    .4|1.88   1.1|B .07   .25| 87.9  87.9| S12  |
```

49	22	33	-.12	.40	1.28	1.7	1.80	1.8	C	.13	.39	63.6	71.2	S49
59	12	33	1.45	.41	1.49	2.3	1.79	2.4	D	.11	.48	57.6	74.6	S59
15	16	33	.81	.39	1.03	.3	1.74	2.4	E	.37	.46	69.7	68.9	S15
30	10	33	1.81	.44	1.54	2.1	1.56	1.6	F	.13	.49	63.6	78.0	S30
22	19	33	.35	.39	1.26	1.8	1.33	1.1	G	.23	.42	54.5	68.4	S22
28	15	33	.96	.40	1.26	1.6	1.29	1.1	H	.28	.46	66.7	70.0	S28
31	25	33	-.65	.44	1.24	1.2	1.28	.7	I	.15	.34	75.8	76.9	S31
38	24	33	-.46	.42	1.26	1.4	1.22	.6	J	.17	.36	66.7	74.6	S38
40	21	33	.04	.40	1.06	.4	1.22	.7	K	.34	.40	69.7	70.0	S40
60	9	33	2.01	.45	1.15	.7	1.21	.7	L	.39	.49	75.8	79.8	S60
54	14	33	1.12	.40	1.18	1.1	1.19	.8	M	.34	.47	63.6	71.4	S54
57	11	33	1.63	.42	1.14	.7	1.15	.6	N	.39	.49	75.8	76.3	S57
41	20	33	.20	.39	1.14	1.0	1.14	.5	O	.31	.41	66.7	68.8	S41
47	8	33	2.22	.47	1.13	.5	1.10	.4	P	.41	.49	78.8	81.7	S47
50	16	33	.81	.39	.96	-.2	1.11	.5	Q	.46	.46	75.8	68.9	S50
4	17	33	.66	.39	1.10	.8	1.06	.3	R	.37	.45	63.6	68.3	S04
18	24	33	-.46	.42	.94	-.2	1.10	.4	S	.37	.36	78.8	74.6	S18
23	15	33	.96	.40	1.04	.3	1.08	.4	T	.43	.46	66.7	70.0	S23
33	17	33	.66	.39	1.07	.5	.96	.0	U	.41	.45	63.6	68.3	S33
37	13	33	1.29	.41	.97	-.1	1.07	.3	V	.48	.48	72.7	73.0	S37
26	27	33	-1.06	.48	1.05	.3	1.04	.3	W	.25	.30	84.8	81.8	S26
2	11	33	1.63	.42	.87	-.6	1.03	.2	X	.55	.49	81.8	76.3	S02
3	5	33	3.00	.56	1.03	.2	.83	-.1	Y	.46	.47	87.9	87.1	S03
6	18	33	.50	.39	1.03	.2	.97	.0	Z	.42	.44	66.7	68.2	S06
	BETTER FITTING OMITTED													
45	31	33	-2.40	.75	.99	.2	.56	-.1		.23	.18	93.9	93.9	S45
24	28	33	-1.31	.51	.98	.0	.73	-.1		.32	.28	84.8	84.8	S24
35	26	33	-.84	.45	.97	-.1	.79	-.2		.36	.32	78.8	79.3	S35
10	16	33	.81	.39	.97	-.2	.88	-.4	z	.49	.46	69.7	68.9	S10
17	27	33	-1.06	.48	.96	.0	.76	-.2	y	.35	.30	84.8	81.8	S17
9	22	33	-.12	.40	.95	-.2	.82	-.3	x	.43	.39	69.7	71.2	S09
13	30	33	-1.94	.63	.95	.1	.67	.0	w	.27	.22	90.9	90.9	S13
7	10	33	1.81	.44	.79	-.9	.95	.0	v	.60	.49	87.9	78.0	S07
44	30	33	-1.94	.63	.93	.0	.64	-.1	u	.29	.22	90.9	90.9	S44
16	25	33	-.65	.44	.92	-.3	.74	-.3	t	.41	.34	75.8	76.9	S16
29	24	33	-.46	.42	.91	-.4	.74	-.4	s	.44	.36	72.7	74.6	S29
36	24	33	-.46	.42	.90	-.5	.71	-.5	r	.45	.36	72.7	74.6	S36
55	31	33	-2.40	.75	.88	.0	.42	-.2	q	.31	.18	93.9	93.9	S55
56	29	33	-1.59	.56	.87	-.2	.63	-.2	p	.36	.25	87.9	87.9	S56
27	21	33	.04	.40	.86	-.9	.76	-.6	o	.51	.40	81.8	70.0	S27
11	28	33	-1.31	.51	.86	-.4	.62	-.3	n	.40	.28	84.8	84.8	S11
53	27	33	-1.06	.48	.85	-.5	.59	-.5	m	.44	.30	78.8	81.8	S53
20	20	33	.20	.39	.84	-1.1	.72	-.8	l	.54	.41	72.7	68.8	S20
42	31	33	-2.40	.75	.84	-.1	.36	-.3	k	.34	.18	93.9	93.9	S42
14	26	33	-.84	.45	.83	-.7	.59	-.6	j	.47	.32	78.8	79.3	S14
25	10	33	1.81	.44	.82	-.7	.82	-.5	i	.61	.49	81.8	78.0	S25
34	21	33	.04	.40	.79	-1.5	.68	-.9	h	.56	.40	81.8	70.0	S34
39	9	33	2.01	.45	.78	-.9	.78	-.5	g	.63	.49	87.9	79.8	S39
58	31	33	-2.40	.75	.78	-.2	.29	-.5	f	.38	.18	93.9	93.9	S58
43	30	33	-1.94	.63	.78	-.4	.39	-.5	e	.42	.22	90.9	90.9	S43
52	29	33	-1.59	.56	.77	-.5	.47	-.4	d	.44	.25	87.9	87.9	S52
21	25	33	-.65	.44	.74	-1.3	.55	-.8	c	.54	.34	81.8	76.9	S21
51	28	33	-1.31	.51	.70	-1.0	.42	-.7	b	.51	.28	84.8	84.8	S51
46	10	33	1.81	.44	.64	-1.7	.56	-1.5	a	.73	.49	87.9	78.0	S46
MEAN	20.5	33.0	.00	.47	.99	.1	.94	.1				77.6	77.9	
S.D.	7.4	.0	1.38	.10	.18	.8	.40	.8				10.0	7.9	

Figure 13: Polarity of ULPPPM Items

The validity of the items is determined based on the fit of each item with the expectations of the Rasch model through statistical analysis of *Infit Mean Square (MNSQ), Outfit MNSQ, Z-standardised (ZSTD),* and *Point Measure Correlation (PT-Mea Corr*). The Rasch model stipulates that ideal fit statistics fall between 0.5 and 1.5, with ZSTD values between -2 and +2 and PT-Mea Corr values exceeding 0.2, indicating that items positively correlate with the measured construct (Bond & Fox, 2015).

The item analysis results for 60 questions indicate that some items meet the expectations of the Rasch model, demonstrating satisfactory construct validity. However, several items were identified as misfit based on fit values exceeding the allowed range or having correlations that were too low.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

Among the items that indicate validity issues are:

i. **S05** (Outfit MNSQ = 2.27, PT-Mea Corr = 0.27)
ii. **S12** (Outfit MNSQ = 1.88, PT-Mea Corr = 0.07)
iii. **S59** (Infit MNSQ = 1.49, PT-Mea Corr = 0.11)
iv. **S30** (Infit MNSQ = 1.54, PT-Mea Corr = 0.13)
v. **S49** (Outfit MNSQ = 1.80, PT-Mea Corr = 0.13)
vi. **S22** (Infit MNSQ = 1.26, PT-Mea Corr = 0.23).

These high values indicate that the responses to these items differ significantly from the model's expectations, thereby suggesting that these items may not consistently measure the assessment literacy construct. Additionally, items such as S12 with a PT-Mea Corr value as low as 0.07 indicate that they do not correlate with the overall ability of the respondents.

Conversely, several items show very good validity, such as S46 (PT-Mea Corr = 0.73), S07 (0.60), and S39 (0.63), all of which demonstrate good fit statistics and high correlation with the measured trait. This indicates that these items are consistent and suitable for measuring the construct of polytechnic lecturers' assessment literacy.

In conclusion, 6 to 7 items were identified as invalid and are recommended for discarding or revision in terms of content formulation. Overall, this instrument demonstrates satisfactory construct validity when measured using the Rasch Model, supported by an item reliability value of 0.87 and an item separation value of 2.64.

Person-Item Map Interpretation of Malaysia Polytechnic Lecturers Assessment Literacy Test Instrument Using the Rasch Model

The Person-Item Map (also known as the Wright Map) displays the distribution of respondent abilities and item difficulties on a shared logit scale**.** Fig. 14 shows the distribution of respondents and items along the logit scale. The mean person is +0.82 logits, indicating that most respondents possess a slightly higher level of assessment literacy ability than the average difficulty of the items, which is centred at 0 logits.

In this study, the logit scale ranged from -3 to +5. The mean item difficulty was centred at 0 logits (a default Rasch setting), while the mean person ability was +0.82 logits. This indicates that respondents, on average, had slightly higher levels of assessment literacy than the average item difficulty.

Items such as S03, S05, S39, and S60 appeared at the top end of the scale (above +2 logits), making them the most difficult. In contrast, items S13, S43, S44, S42, S45, S55, and S58 appeared below -2 logits, indicating they were the easiest. The most able respondent (ID: 21324131) was positioned around +5 logits, while the least able were situated below -2 logits. According to Linacre (2011), such gaps between person and item distributions may indicate the need for additional items, especially at the upper end, to better capture higher abilities.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1084-1109

```
                              Person - MAP - Item
                                  <more>|<rare>
 5                                      +
                                        |
                                        |
                                        |
                                        |
                                        |
                              21324131  |
 4                                      +
                                        |
                                        |
                                        |
                              21224142  |
                                       T|
                              11114131  |
 3                                      +  S03
                                        |
                              11324131  |T S05
                                        |
                                        |
                              21112131  |  S47
                                        |
 2                                     S+  S39   S60
                              11324132  |  S07   S25   S30   S46
                              21234133  |
                              21324142  |  S02   S57
                                        |S S59
                                        |  S37
                     21214134 21324144  |  S54
 1                            21314230  +  S08   S23   S28
            21112131 22114133 22224131 M|  S10   S15   S50
            12234132 21332131 22112131  |  S04   S33
                                        |  S06
            11124132 11324132 11324132  |
                              11324134  |  S19   S22
   11224133 12314134 21122143 21224132  |  S20   S41
 0                                      +M S27   S34   S40   S48
                     11314141 21324133  |  S09   S49
                              21314131  |
                     21122131 21324143 S|  S01   S18   S29   S36   S38
                              21224131  |
                                        |  S16   S21   S31
                                        |  S14   S35
-1                            11214121  +  S17   S26   S53
                              12224232  |
                                        |  S11   S24   S32   S51
                                        |S
                                       T|  S12   S52   S56
                                        |
                                        |
-2                                      +  S13   S43   S44
                                        |
                                        |
                                        |  S42   S45   S55   S58
                                        |
                                        |T
                                        |
-3                                      +
                                  <less>|<frequ>
```

Figure 14: Person-Item Map Interpretation

The Rasch analysis confirms that the UL3PM instrument has strong psychometric characteristics. The person-item map shows a generally well-aligned match between item difficulty and respondent ability. Person and item reliability are both above 0.87, while separation indices indicate good classification power. Nevertheless, several misfitting items (e.g., S05, S59, S30) were detected and should be revised or removed. Additionally, several item pairs (e.g., S43–S55) showed high residual correlations, suggesting local item dependency. Addressing these issues will enhance the construct validity, unidimensionality, and precision of the instrument. In summary, the instrument is suitable for use in measuring assessment literacy among polytechnic lecturers, pending minor adjustments to misfitting and redundant items.

Volume 11 Issue 64 (September 2026) PP. 1084-1109

Conclusion

In summary, the UL3PM is suitable for use in measuring assessment literacy among polytechnic lecturers, pending minor adjustments to misfitting and redundant items. The Rasch analysis confirms that the UL3PM instrument has strong psychometric characteristics. It had high item and individual reliability with values of 0.87 and 0.91 respectively. The Item Separation Index = 2.64, indicating that the number of respondents is sufficient to distinguish items well by difficulty level. The Person Separation value is 3.11, which allows lecturers to be classified into at least 4 different ability levels. The Rasch model demonstrates satisfactory construct validity. However, several items were identified as misfits because their fit values exceeded the allowed range or their correlations were too low. Items 6 to 7 were identified as invalid and are recommended for discarding or revision in terms of content formulation. The person-item map shows a generally well-aligned match between item difficulty and respondent ability. Person and item reliability are both above 0.87, while separation indices indicate good classification power. Nevertheless, several misfitting items (e.g., S05, S59, S30) were detected and should be revised or removed. Additionally, several item pairs (e.g., S43–S55) showed high residual correlations, suggesting local item dependency. Addressing these issues will enhance the construct validity, unidimensionality, and precision of the instrument.

Acknowledgements: The authors would like to express their sincere gratitude to the Bahagian Kompetensi dan Peningkatan Kerjaya and the Examination and Evaluation Division at Jabatan Pendidikan Politeknik dan Kolej Komuniti (JPPKK), and to Politeknik Tun Syed Nasir Syed Ismail (PTSN) lecturers who assisted with data collection and instrument validation.

Finally, the authors would like to thank and express their gratitude to the Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia (UTHM) for its continuous cademic guidance, institutional support, and encouragement throughout the completion of this research.

Funding Statement: The author also extends their heartfelt appreciation to Bahagian Kompetensi dan Peningkatan Kerjaya for providing funding to enable the presentation of this research paper at INCEDD 2026 in Vietnam.

Conflict of Interest Statement: The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Education, Psychology and Counselling (IJEPC)

Ethics Statement: This study was conducted in accordance with ethical research standards. All procedures involving human participants were reviewed and approved by the Jabatan Pendidikan Politeknik dan Kolej Komuniti (JPPKK), approval number JPPKK/PPI/700-1/1 Jld 35(13). Informed consent was obtained from all participants prior to data collection. Participation was voluntary, and respondents were assured of confidentiality and anonymity. The data collected were used solely for academic purposes.

Author Contribution Statement: All authors contributed significantly to the development of this manuscript. Normazita Mat Ali was responsible for the conceptualization, methodology, and overall supervision of the study. Nurfirdawati Muhamad Hanafi handled data collection, analysis, and interpretation of results. Razif Salim contributed to the literature review, drafting, and critical revision of the manuscript. All authors read and approved the final version of the manuscript prior to submission.

References

Afshar, H. S., & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, *70*, 101042.

Ahn, I., & Bong, M. (2020). Self-efficacy and self-concept as predictors of teacher-student relationships and academic achievement: A longitudinal study. *Journal of Educational Psychology, 112*(4), 715–729. https://doi.org/10.1037/edu0000385

Anggara, D. S., & Abdillah, C. (2023). Content validity analysis of literacy assessment instruments. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, *42*(2), 447-459.

Aria, D., Sukyadi, D., & Kurniawan, E. (2021). Teacher assessment literacy: Indonesian EFL secondary teachers' self-perceived on classroom-based assessment practice. *English Review: Journal of English Education*, *10*(1), 15-26.

Arjana, I. M., Parmiti, D. P., Candiasa, I. M., & Widiartini, N. K. (2025). Unlocking the predictive power of the Rasch model: A systematic literature review on educational instrument calibration and assessment accuracy. *Edelweiss Applied Science and Technology, 9*(6), 2003–2022.

Asres, A. M., & Abebe, A. S. (2024). Teachers' Competence in The Educational Assessment of Students in A Competence Based Education in Amhara Region, Ethiopia: The Case of TVET Teachers and Assessors.

Atjonen, P., Pöntinen, S., Kontkanen, S., and Ruotsalainen, P. (2022). In enhancing preservice teachers' assessment literacy: focus on knowledge base, conceptions of assessment, and teacher learning. *Front. Educ.* 7, 891391. doi: 10.3389/feduc.2022.891391

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway.

Bond, T.G. and C.M. Fox, 2007. Applying the Rasch Model. Fundamental Measurement in the Human Sciences. 2nd Edn., Lawrence Erlbaum Associates Publishers, New Jersey, ISBN: 0805854622.

Bond, T.G. and C.M. Fox, 2007. Applying the Rasch Model. Fundamental Measurement in the Human Sciences. 2nd Edn., Lawrence Erlbaum Associates Publishers, New Jersey, ISBN: 0805854622.

Boone, W. J. (2021). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education, 20*(1), rm1. https://doi.org/10.1187/cbe.20-04-0076

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices*(3rded.).PearsonEducation. https://www.academia.edu/39088154/Language_Assessment_Principles_and_Classroom_Practices_H_Douglas_Brown_H_Douglas_Brown_Priyanvada_Abeywickrama

Chong, J., Mokshein, S. E., & Mustapha, R. (2021). A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VOTAL). *International Journal of Academic Research in Business and Social Sciences*, *11*(4). https://doi.org/10.6007/ijarbss/v11-i4/9547

Christoforidou, M., and Kyriakides, L. (2021). Developing teacher assessment skills: the impact of the dynamic approach to teacher professional development. *Stud. Educ. Eval.* 70, 101051. doi: 10.1016/j.stueduc.2021.101051

Diao, J., & Hu, K. (2022). Preparing TVET teachers for sustainable development in the information age: Development and application of the TVET teachers' teaching competency scale. *Sustainability*, *14*(18), 11361.

Diao, J., Han, X., Zhou, Q., & Wang, Y. (2023). Professional Competencies in TVET: Framework, Indicators and Assessment Instrument. In *Handbook of Technical and*

Vocational Teacher Professional Development in the Digital Age (pp. 41-73). Singapore: Springer Nature Singapore.

Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, *12*(1), 124-134.

Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, *19*(5), 503-525.

Giraldo, F. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile Issues in TeachersProfessional Development*, *23*(1), 197-213.

Giraldo, F. (2021a). Language assessment literacy and teachers' professional development: A review of the literature. *Profile Issues in TeachersProfessional Development*, *23*(2), 265-279.

Hamzah, F. M., Rashid, M. N. A., Rahman, M. N. A., & Rasul, M. S. (2022). Evaluating the validity and reliability of authentic learning instruments using Rasch Model: A study in Malaysian polytechnics. *International Journal of Global Optimization and Its Application,* *1*(3), 182–189. https://myjms.mohe.gov.my/index.php/ijgoia/article/view/20606

Hannigan, C., Alonzo, D., & Oo, C. Z. (2022). Student assessment literacy: Indicators and domains from the literature. *Assessment in Education: Principles, Policy & Practice*, *29*(4), 482-504.

Hashim, S., Zakariah, S. H., Taufek, F. A., Zulkifli, N. N., Lah, N. H. C., & Murniati, D. E. (2021). An observation on implementation of classroom assessment in technical and vocational education and training (TVET) subject area. *Journal of Technical Education and Training*, *13*(3), 190-200.

Herppich, S., Praetorius, A.-K., Forster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teach. Teach. Educ.* 76, 181–193. doi: 10.1016/j.tate.2017.12.001

Ismail, N., Abdullah, S., & Othman, N. (2022). Assessment literacy among technical and vocational education and training (TVET) lecturers in Malaysia: Challenges and prospects. *Journal of Technical Education and Training,* *14*(3), 12–25. https://doi.org/10.30880/jtet.2022.14.03.002

Ismail, N., Abdullah, S., & Othman, N. (2022). Assessment literacy among technical and vocational education and training (TVET) lecturers in Malaysia: Challenges and prospects. *Journal of Technical Education and Training,* *14*(3), 12–25. https://doi.org/10.30880/jtet.2022.14.03.002

Istiara, F., & Hastomo, T. (2023). Exploring lecturers and administrative staffs' strategies to hone EFL students' digital literacy. *JOALL (Journal of Applied Linguistics and Literature)*, *8*(1), 151-172.
Fauzan, F., Arifin, F., Lubis, M. A., & Firdaus, F. M. (2022). Lecturer's Digital Literacy Ability in the Pandemic. *Cypriot Journal of Educational Sciences*, *17*(4), 1130-1142.

Jaya, I., & Nasir, M. (2025). Holistic Assessment in TVET: Enhancing Writing Evaluation Through a Pilot Study. *Journal of Language Testing and Studies*, *1*(1), 55-68.

Jean de Dieu, H., & Zhou, K. (2021). Physical literacy assessment tools: a systematic literature review for why, what, who, and how. *International Journal of Environmental Research and Public Health*, *18*(15), 7954.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Juanjuan, G., & Mohd Yusoff, N. (2022). A review of teachers' assessment literacy proficiency measures. International Journal of Evaluation and Research in Education (IJERE), 11(3), 1143. DOI: 10.11591/ijere.v11i3.22300

Kalajahi, S. a. R., & Abdullah, A. N. (2016). Assessing Assessment Literacy and Practices among Lecturers. *Pedagogika*, *124*(4), 232–248. https://doi.org/10.15823/p.2016.65

Kao, Y. T. (2023). Empowering preservice English teachers with language assessment literacy concepts and practices: Application of Vygotskian concept-based language instruction. *The Modern Language Journal*, *107*(S1), 68-94.

Khalid, N. H. M., Latif, A. A., & Yusof, I. J. (2021). Research Focusing on Assessment Literacy: A Systematic Literature review. *Studies of Applied Economics*, *39*(10). https://doi.org/10.25115/eea.v39i10.5816

Khalid, N. H. M., Yusof, I. J., Latif, A. A., & Jani, M. D. M. (2023). Psychometric properties of teacher classroom assessment literacy instrument using Rasch model analysis. *Int J Eval & Res Educ ISSN*, *2252*(8822), 8822.

Khalid, N. H. M., Yusof, I. J., Latif, A. A., & Jani, M. D. M. (2023). Psychometric properties of teacher classroom assessment literacy instrument using Rasch model analysis. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(2), 638. https://doi.org/10.11591/ijere.v12i2.23607

Khine, M. S. (2020). Rasch measurement. In *Springer eBooks*. https://doi.org/10.1007/978-981-15-1800-3

Linacre, J. M. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions, 7*(4), 328.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Linacre, J. M. (2003). *Winsteps computer program version 3.48. Chicago*. http://www.winsteps.com/a/ winsteps-tutorial

Linacre, J. M. (2005). *Winstep Rasch measurement computer program*. MESA Press.

Linacre, J. M. (2007). *A user's guide to WINSTEPS Rasch-model computer program*. MESA Press.

Linacre, J. M. (2010). When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions, 23*(4), 1241.

Linacre, J. M. (2012). *A user's guide to WINSTEPS: Rasch model computer programs*. MESA Press.

Linacre, J. M. (2017). Winsteps Rasch measurement computer program user's guide. *Winsteps.com*.

Linacre, J.M., 2011. Winsteps Rasch Measurement Computer Program Users Guide. Beaverton Publisher, Beaverton, Oregon,.

Lutovac, S., & Flores, M. A. (2022). Conceptions of assessment in pre-service teachers' narratives of students' failure. *Cambridge Journal of Education*, *52*(1), 55-71.

Mardiana, H. (2024). Perceived impact of lecturers' digital literacy skills in higher education institutions. *Sage Open*, *14*(3), 21582440241256937.

Modarresi, G., Jalilzadeh, K., Coombe, C., & Nooshab, A. (2021). Validating a test to measure translation teachers' assessment literacy. *Journal of Asia TEFL*, *18*(4), 1503.

Mokshein, S. E., Ahmad, H., Lebar, O., Dollah, M. U., Yunus, J., Rahmat, A., & Ahmed, H. H. (2022). Validation of the Malaysian-Based Assessment Practice Inventory for Teacher

Educators (MAPITE) using Rasch Model. Journal of Engineering and Applied Sciences, 17(9), 2783–2798. doi:10.36478/jeasci.2019.2783.2798

Mudau, P. K. (2022). Lecturers' views on the functionality of e-portfolio as alternative assessment in an open distance e-learning. *International Journal of Educational Methodology*, *8*(1), 81-90.

Muhammad Hanafi, N. B., Mat Ali, N. B., Zamani, S. B., Yamin, N. A. B., & Ismail, N. B. (2020). Examining assessment literacy: A study of technical teachers. *European Journal of Molecular & Clinical Medicine, 7*(8), 705–715.

Mutinda, F. M. (2023). *Classroom Assessment Practices by Mathematics Teachers in Technical Vocational Education Training (TVET) Institutions in Kenya* (Doctoral dissertation, University of Nairobi).

Njuki, J. W., & Mukundi, R. M. (2023). Enhancing integration of Information Literacy in TVET curriculum in response to COVID-19 pandemic effects. *American Journal of Education and Practice,[Lewes, DE]*, *7*(4), 42-53.

Norhasim, N., & Mohamad, M. (2020). *Teachers' challenges on the implementation of classroom-based assessment for oral proficiency: A literature review*. *Creative Education, 11*(11), 1922–1929. https://doi.org/10.4236/ce.2020.1110140

Nur, L., Nurhasan, N., Al Ardha, M. A., Chaeroni, A., Friskawati, G. F., Muslihin, H. Y., ... & Malik, A. A. (2025). Measurement of physical literacy of elementary school students in Indonesia: an analysis using the Rasch model. *Retos: nuevas tendencias en educación física, deporte y recreación*, (65), 909-920.

Oo, C. Z., Alonzo, D., & Asih, R. (2022). Acquisition of teacher assessment literacy by pre-service teachers: A review of practices and program designs. *Issues in Educational Research*, *32*(1), 352-373.

Osman, K., Sovey, S., & Matore, M. E. E. M. (2022). RasCh Analysis for Disposition Levels of Computational Thinking Instrument among Secondary school students. *Eurasia Journal of Mathematics Science and Technology Education*, *18*(3), em2088. https://doi.org/10.29333/ejmste/11794

Padmadewi, N. N., Artini, L. P., Ratminingsih, N. M., & Marsakawati, N. P. E. (2022). Needs Analysis of Literacy Assessment Using Blended Learning for Beginner EFL Learners. *Journal of Language Teaching & Research*, *13*(2).

Pastore, S. (2023). Teacher assessment literacy: a systematic review. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1217167

Purnami, D. E., Hartinah, S., & Susongko, P. (2024). Model Assesment Literasi Membaca pada Siswa Sekolah Dasar Berbasis Pemodelan Rasch. *Journal of Education Research*, *5*(3), 3217-3223.

Ramaligela, S. M. (2022). Student's Perceptions of assessment legitimacy in TVET Sector in South Africa. In *Research Anthology on Vocational Education and Preparing Future Workers* (pp. 452-467). IGI Global Scientific Publishing.

Rasch, G. (1980). Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: University of Chicago Press.

Rasul, M. S., Ismail, N., Rajuddin, M. R., & Rauf, R. A. A. (2009). Aspects of employability skills in technical and vocational education and training (TVET) curriculum. *Journal of Technical Education and Training, 1*(1), 1–8. https://publisher.uthm.edu.my/ojs/index.php/jtet/article/view/529

Rohaya Talib. (2009). Pembinaan dan Pengesahan Instrumen Ujian Literasi Pentaksiran. Tesis PhD, Universiti Teknologi Malaysia.

Rozali, M. Z., Puteh, S., Yunus, F. A. N., Hamdan, N. H., & Latif, H. F. M. (2022). Reliability and validity of instrument on Academic Enhancement Support for Malaysian student-athletes using Rasch Measurement Model. *Asian Journal of University Education, 18*(1), 1–15. https://doi.org/10.24191/ajue.v18i1.17199

Singh, H. K. J., Singh, C. K. S., Ong, E. T., Wong, W. L., & Gopal, R. et al. (2023). *Developing items to measure the assessment literacy of ESL teachers*. *Journal of Higher Education Theory and Practice, 23*(16), 225–245.

Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, *74*, 101157.

Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, *39*(4), 649-678.

Tsagari, D. (2021). Gauging The Assessment Literacy Levels Of English Language Teachers In Norway. *The European Journal of Applied Linguistics and TEFL*, *10*(1), 161-192.

Vinodhen, V. (2025). Impact of online resources on assessment literacy of polytechnic educators. *Forum for Education Studies.*, *3*(3), 2678. https://doi.org/10.59400/fes2678

Wright, B. D., & Master, G. N. (1982). *Rating scale analysis*. Cesa Press.

Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. In E. V. Smith, & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models, and applications*. JAM Press.

Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. In E. V. Smith, & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models, and applications*. JAM Press.

Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation*, *74*, 101183.

Yusop, S. R. M., Rasul, M. S., Mohamad Yasin, R., Hashim, H. U., & Jalaludin, N. A. (2022). An assessment approaches and learning outcomes in technical and vocational education: A systematic review using PRISMA. *Sustainability, 14*(9), 5225. https://doi.org/10.3390/su14095225

Yusop, S. R. M., Rasul, M. S., Yasin, R. M., & Hashim, H. U. (2023). Identifying and validating vocational skills domains and indicators in classroom assessment practices in TVET. *Sustainability*, *15*(6), 5195. https://doi.org/10.3390/su15065195